



**Titre:** Définition d'un modèle d'optimisation pour le dimensionnement de  
Title: réseaux troisième génération

**Auteur:** Catherine Voisin  
Author:

**Date:** 2002

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Voisin, C. (2002). Définition d'un modèle d'optimisation pour le dimensionnement  
Citation: de réseaux troisième génération [Mémoire de maîtrise, École Polytechnique de  
Montréal]. PolyPublie. <https://publications.polymtl.ca/7017/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/7017/>  
PolyPublie URL:

**Directeurs de  
recherche:** Brigitte Jaumard, & Richard Gourdeau  
Advisors:

**Programme:** Non spécifié  
Program:

UNIVERSITÉ DE MONTRÉAL

DÉFINITION D'UN MODÈLE D'OPTIMISATION POUR LE  
DIMENSIONNEMENT DE RÉSEAUX TROISIÈME GÉNÉRATION

CATHERINE VOISIN  
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES (M.Sc.A.)  
(TÉLÉCOMMUNICATIONS)  
DÉCEMBRE 2002



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-81529-3

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

DÉFINITION D'UN MODÈLE D'OPTIMISATION POUR LE  
DIMENSIONNEMENT DE RÉSEAUX TROISIÈME GÉNÉRATION

présenté par: VOISIN Catherine

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

M. HERTZ Alain, Ph.D., président

Mme. JAUMARD Brigitte, Ph.D., Th. d'Hab., membre et directeur de recherche

M. GOURDEAU Richard, Ph.D., membre et codirecteur de recherche

M. GIRARD André, Ph.D., membre



À mes parents,  
merci pour votre soutien continu et votre enthousiasme.

## PRÉFACE

Ce projet de recherche a été très enrichissant et motivant car il m'a permis de travailler sur la conceptualisation, le développement et la validation d'un outil de dimensionnement pour les réseaux de troisième génération. Les différentes étapes du projet ont été les suivantes :

- la modélisation du réseau et l'élaboration de la stratégie de dimensionnement (Mars-Septembre 2001),
- la mise en équation du modèle (Septembre-Décembre 2001),
- le développement et la programmation (Janvier-Juillet 2002),
- la validation et l'exploitation des résultats (Août-Octobre 2002).

J'ai eu la chance d'être accompagnée par Brigitte Jaumard, ma directrice de recherche, tout au long de ce projet ainsi que Daniel Dufour et Francis Lupien d'Ericsson *Research Canada*, ce dernier m'ayant particulièrement aidée lors de la phase de développement à intégrer des valeurs réalistes pour les paramètres du modèle. Je tiens, en-outre, à souligner l'aide de Rolland Malhamé, Professeur à l'Ecole Polytechnique (et mon co-directeur de recherche de août 2001 à septembre 2002), qui a participé aux discussions dans cette étape difficile qu'était la modélisation et l'élaboration de la stratégie de dimensionnement.

Comme pour tout travail de modélisation, les discussions ont été nombreuses, toujours très animées, couvrant parfois des sujets très larges et parfois des sujets très ciblés. Certaines idées ont été approfondies, certaines abandonnées et d'autres mises en attente. En conséquence, le modèle présenté dans ce mémoire ne reflète qu'une partie des efforts qui ont été consacrés à ce sujet. Je tiens cependant à indiquer que Monsieur Malhamé a particulièrement porté son attention sur les formules de capacité radio adaptées au trafic multi-services, de plus, il nous a éclairé sur l'aspect anticipatif du contrôle d'admission modélisé et sur le fait que le dimensionnement obtenu avec notre modèle fournit une borne supérieure de performance par rapport aux dimensionnements basés sur les contrôles d'admission causaux, implantés dans la réalité. Enfin, il nous a permis de conceptualiser les contraintes sur les débits des liens câblés pour garantir les délais en utilisant le modèle de Parekh et

Gallager<sup>1, 2</sup>. Nos discussions ont aussi porté sur le dimensionnement des mémoires tampons au niveau des routeurs. Cet aspect, non intégré dans la version finale du modèle présentée dans ce mémoire, est cependant développé dans le rapport intermédiaire<sup>3</sup> rédigé dans le cadre du contrat de recherche avec Ericsson.

Finalement, malgré la difficulté des concepts développés au cours de ce travail, Madame Jaumard a permis de maintenir une ligne directrice pour assurer la réalisabilité du modèle en tenant compte des contraintes liées à une procédure d'optimisation.

---

1. A K Parekh et R G Gallager, A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case, *IEEE/ACM Transactions on Networking* 1(3), 1993, 344-357.

2. A K Parekh et R G Gallager, A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple-Node Case, *IEEE/ACM Transactions on Networking* 2(2), 1994, 137-150.

3. B Jaumard et R Malhamé et C Voisin, Deliverable #2: Definition of an Optimization Model for the Dimensioning of 3G Multimedia Networks, *École Polytechnique de Montréal*, février 2002.

## REMERCIEMENTS

Je tiens à remercier ma directrice de recherche, Brigitte Jaumard, pour la qualité de son encadrement, son aide financière et son soutien au cours de ce projet aux multiples rebondissements.

Ces remerciements s'adressent également à Monsieur Roland Malhamé pour ses remarques constructives au début de ma maîtrise.

Je tiens, en outre, à souligner ma reconnaissance à Francis Lupien et Daniel Dufour d'Ericsson Research Canada (*LMC*) pour leur expertise d'ingénieurs qui s'est avérée précieuse sur ce sujet à forte connotation industrielle.

Je n'oublierai pas mon école d'origine, Supélec, (*l'École Supérieure d'Electricité*), qui m'a permis, grâce à son programme de formation à l'étranger, de profiter d'une expérience humaine et professionnelle très enrichissante.

Je voudrais citer mon Stéphane qui est resté dans mon cœur et dans mes pensées pendant ces deux années outre-atlantique, nos excursions perdues dans les grands espaces canadiens sont un souvenir inoubliable. Je n'oublierai pas Sandrine, amie depuis Supélec, nous nous sommes lancées ensemble à la découverte du Québec et c'est une belle expérience que je suis heureuse d'avoir partagée avec elle.

Je remercie aussi Claudia, Geneviève et Bernard pour leur gentillesse, j'ai eu la chance de partager leur quotidien très animé pendant une année. Je finirai en remerciant mon collègue de bureau Stéphane ainsi que Sébastien pour son interface Interplex qui m'a bien aidée et toute la gang du GERAD pour leurs conseils avisés, pour l'entre-aide et la chaleureuse ambiance dont ils sont tous les acteurs.

## RÉSUMÉ

Le défi des futurs réseaux de communication mobile troisième génération est de permettre aux usagers d'accéder à toutes sortes d'application depuis leurs terminaux mobiles. Les applications multimédia, en particulier, ont des exigences de haut débit et de qualité de service que ces futurs réseaux devront satisfaire. Les nouveaux profils de trafic, très gloutons, vont nécessiter de nombreuses ressources. Ainsi, les outils de dimensionnement joueront un rôle primordial pour les fournisseurs d'équipements et de services, l'objectif étant, pour ces outils de dimensionnement, de minimiser les ressources nécessaires tout en satisfaisant la demande de trafic.

Dans ce mémoire, nous proposons de développer un modèle d'optimisation pour le dimensionnement de ces futurs réseaux de communication mobile. Une première étape consiste à modéliser ces réseaux en intégrant les mécanismes qui influencent le dimensionnement. Ensuite, la stratégie de dimensionnement est définie, l'objectif du problème étant de minimiser le nombre de stations de base et les capacités des liens câblés à déployer.

La formulation mathématique de la modélisation conduit à un problème linéaire mixte. CPLEX-MIP et ses stratégies de branchement sont utilisés pour résoudre des jeux de données et donnent des solutions optimales ou heuristiques suivant la taille des problèmes. Une fois le modèle validé du point de vue du dimensionnement, nous proposons d'étudier l'évolution du dimensionnement optimal en fonction du trafic escompté pour un réseau test.

## ABSTRACT

The challenge of future third generation communication wireless networks is to let users to access multiple kinds of application from their mobile stations. Particularly, the multimedia applications have strong requirements on throughput and quality of service that the future networks must satisfy. The new traffic profiles, very greedy, will require a lot of ressources. Thus, dimensionning tools will play a crucial role for the equipment and services providers. The aim of these dimensionning tools is to minimize the ressources needed to satisfy the traffic demand.

In this thesis, we propose to developp an optimization model for the dimensioning of these futur wireless communication networks. A first step consists in modeling these networks by integrating the mechanisms that influence the dimensioning. Then, the dimensioning strategy is defined, the objective being to minimize the number of base stations and the capacities of the wired links to deploy.

The mathematical formulation of the model leads to a linear and mixed problem. CPLEX-MIP and its branching strategies are used to solve problems with optimal solutions or heuristic solutions depending on their sizes. Once the model validated for the dimensioning, we propose to study the evolution of the optimal design for several traffic instances and for a given network architecture.

## TABLE DES MATIÈRES

<b>DÉDICACE . . . . .</b>	<b>iv</b>
<b>PRÉFACE . . . . .</b>	<b>v</b>
<b>REMERCIEMENTS . . . . .</b>	<b>vii</b>
<b>RÉSUMÉ . . . . .</b>	<b>viii</b>
<b>ABSTRACT . . . . .</b>	<b>ix</b>
<b>TABLE DES MATIÈRES . . . . .</b>	<b>x</b>
<b>LISTE DES TABLEAUX . . . . .</b>	<b>xiv</b>
<b>LISTE DES FIGURES . . . . .</b>	<b>xvi</b>
<b>LISTE DES NOTATIONS ET DES SYMBOLES . . . . .</b>	<b>xviii</b>
<b>LISTE DES ANNEXES . . . . .</b>	<b>xx</b>
<b>INTRODUCTION . . . . .</b>	<b>xxi</b>
0.1 Motivations . . . . .	xxi
0.2 Organisation du mémoire . . . . .	xxii
0.3 Contributions . . . . .	xxiii
<b>CHAPITRE 1: ARCHITECTURE D'UN RÉSEAU D'ACCÈS CDMA2000</b>	
<b>ET SERVICES MULTIMÉDIA . . . . .</b>	<b>1</b>
1.1 Les réseaux troisième génération (3G) . . . . .	1
1.1.1 Historique . . . . .	1
1.1.2 CDMA2000 versus UMTS . . . . .	2
1.2 Architecture d'un réseau d'accès CDMA-2000 . . . . .	4
1.2.1 Le réseau d'accès radio (RAN) . . . . .	5
1.2.2 Le réseau de cœur . . . . .	6

1.3	Les principaux mécanismes du réseau d'accès radio . . . . .	7
1.3.1	Le contrôle de puissance . . . . .	7
1.3.2	Le handover . . . . .	8
1.4	La qualité de service dans les réseaux 3G . . . . .	9
1.4.1	Les classes de service . . . . .	10
1.4.2	Transport des applications temps réel . . . . .	11
1.4.3	Le service intégré ou comment fournir la qualité de service . . . . .	12
1.5	Conclusions . . . . .	19

## **CHAPITRE 2: REVUE DE LA LITTÉRATURE SUR LES MODÈLES D'OPTIMISATION DÉJÀ EXISTANTS . . . . . 21**

2.1	Dimensionnement des réseaux d'accès radio . . . . .	21
2.2	Dimensionnement des réseaux câblés . . . . .	25
2.3	Conclusions . . . . .	29

## **CHAPITRE 3: MODÉLISATION ET STRATÉGIE DE DIMENSIONNEMENT . . . . . 30**

3.1	Modélisation du réseau . . . . .	31
3.1.1	Architecture . . . . .	32
3.1.2	Politique d'ordonnancement des paquets sur les liens câblés . . . . .	32
3.1.3	Contrôle d'admission de bout en bout . . . . .	36
3.1.4	Le handover . . . . .	38
3.1.5	Formules de capacité radio . . . . .	40
3.2	Modélisation du trafic . . . . .	49
3.2.1	Les applications . . . . .	49
3.2.2	Modélisation d'une session . . . . .	50
3.2.3	Routage . . . . .	54
3.2.4	Durée de la planification et séquençement temporel . . . . .	54
3.2.5	Délai . . . . .	55
3.3	Limites du modèle . . . . .	62
3.3.1	Modèle anticipatif . . . . .	62
3.3.2	Imprécisions sur le dimensionnement dues au séquençement temporel . . . . .	62



3.3.3	Les formules de capacités radio . . . . .	63
3.3.4	Le délai pour les sessions en <i>soft</i> handover . . . . .	63
3.4	Stratégie de dimensionnement . . . . .	64
3.5	Limites de la stratégie . . . . .	65
3.5.1	Distribution uniforme . . . . .	65
3.5.2	Les BSCs et les stations de base potentielles pour un usager mobile .	66
3.6	Récapitulatif des points principaux de la modélisation . . . . .	67

## CHAPITRE 4: FORMULATION MATHÉMATIQUE DU MODÈLE D'OPTIMISATION . . . . . 70

4.1	Notations . . . . .	70
4.1.1	Notations générales . . . . .	70
4.1.2	Notations pour les niveaux de service et la qualité de service . . . .	71
4.1.3	Notations pour les sessions . . . . .	71
4.1.4	Notations pour les liens câblés . . . . .	73
4.1.5	Notations pour les liens radio . . . . .	73
4.2	Définition des variables . . . . .	73
4.3	Fonction objectif . . . . .	76
4.4	Contraintes . . . . .	76
4.4.1	Contrôle d'admission . . . . .	76
4.4.2	Niveaux de service . . . . .	78
4.4.3	Liaison entre la partie câblée et la partie radio . . . . .	78
4.4.4	Qualité de service: délai . . . . .	79
4.4.5	Liaison entre le réseau de coeur et le réseau d'accès radio . . . . .	82
4.4.6	Qualité de service: sélection du RAB sur les liens radio . . . . .	82
4.4.7	<i>soft</i> handover . . . . .	84
4.4.8	Capacités des liens câblés . . . . .	85
4.4.9	Capacités des liens radio . . . . .	86
4.4.10	Activation des stations de base potentielles . . . . .	89
4.5	Récapitulatif du modèle d'optimisation . . . . .	89
4.5.1	Définition et bornes des variables . . . . .	89

4.5.2	Formulation mathématique du modèle . . . . .	90
<b>CHAPITRE 5: RÉSULTATS . . . . .</b>		<b>93</b>
5.1	Implémentation . . . . .	93
5.1.1	Générateur de trafic . . . . .	93
5.1.2	Valeurs des paramètres dans les contraintes . . . . .	99
5.1.3	Procédure d'optimisation : CPLEX . . . . .	104
5.2	Description des jeux de données . . . . .	108
5.2.1	Topologies tests des réseaux de cœur et des BSCs . . . . .	109
5.2.2	Profil de trafic . . . . .	109
5.2.3	Nombre de stations de base potentielles . . . . .	110
5.2.4	Description des jeux de données . . . . .	111
5.3	Stratégies pour la résolution des problèmes . . . . .	113
5.4	Validation du modèle . . . . .	116
5.4.1	Validation de la contrainte qui relie les débits des flots d'une session au niveau de la transition entre le réseau de cœur et le réseau d'accès radio . . . . .	116
5.4.2	Influence du coût des stations de base sur le dimensionnement . . . .	117
5.4.3	Impact du nombre de stations de base potentielles sur le dimension- nement . . . . .	120
5.4.4	Optimalité des positions des stations de base . . . . .	121
5.4.5	Le <i>soft</i> handover . . . . .	125
5.5	Résultats pour le dimensionnement des réseaux 3G . . . . .	127
5.5.1	Taux d'occupation des stations de base . . . . .	127
5.5.2	La qualité de service et les niveaux de service par application . . . .	130
5.5.3	Dimensionnement des réseaux 3G . . . . .	135
5.5.4	Conclusions . . . . .	137
<b>CONCLUSION . . . . .</b>		<b>140</b>
<b>RÉFÉRENCES . . . . .</b>		<b>144</b>

## LISTE DES TABLEAUX

1.1	Différences techniques entre UMTS et CDMA-2000 . . . . .	3
1.2	Les 4 classes de services définies par 3GPP . . . . .	11
1.3	Avantages et inconvénients des principales politiques d'ordonnancement de paquets . . . . .	18
3.1	Configuration radio RC4 . . . . .	40
3.2	Les applications du modèle . . . . .	49
3.3	Valeur de $\mathcal{R}(r_t)$ en fonction du débit sélectionné sur le lien radio descendant	53
3.4	Exigences sur le délai par application . . . . .	57
3.5	Délai moyen de transmission d'une trame sur un lien radio . . . . .	61
5.1	Profils d'utilisation des applications par les usagers . . . . .	96
5.2	Attributs d'une session suivant son application . . . . .	96
5.3	Débits déclarés pour les flots des sessions CBR . . . . .	100
5.4	Bornes sur les débits pour les flots des sessions VBR . . . . .	100
5.5	Valeurs des taux de blocage $Br_{ap}$ . . . . .	100
5.6	Voix . . . . .	101
5.7	Vidéo conférence . . . . .	101
5.8	Flux vidéo . . . . .	101
5.9	Navigation Internet . . . . .	102
5.10	Courrier électronique . . . . .	102
5.11	Paramètres pour le délai d'une session de navigation Internet . . . . .	104
5.12	Paramètres pour le délai d'une session de courrier électronique . . . . .	104
5.13	Instances (01)1-8, (01)1-30, (01)2-16 . . . . .	111
5.14	Instances (01)3-16 , (01)4-20 et (02)5-25 . . . . .	112
5.15	Combinaisons des stratégies de branchement . . . . .	115
5.16	Performances des combinaisons de stratégies de branchement . . . . .	115
5.17	Influence du coût des stations de base . . . . .	118

5.18 Impact du nombre de stations de base potentielles sur le dimensionnement pour le profil de trafic 1 (60 sessions) . . . . .	120
5.19 Configurations optimales pour le profil de trafic 1 (60 sessions) . . . . .	121
5.20 Combinaisons des stations de base activées à partir d'une même configuration initiale . . . . .	123
5.21 Taux d'utilisateurs en <i>soft</i> handover . . . . .	125
5.22 Répartition des sessions selon les applications pour les profils de trafic P1, P2 et P3 . . . . .	128
5.23 Navigation Internet High . . . . .	131
5.24 Répartition des sessions selon les niveaux de qualité de service $Q_{r,a,p}$ pour le profil P2 (80 sessions) . . . . .	132
5.25 Répartition des sessions acceptées pour l'instance associée au profil P2 . . . . .	132
5.26 Taux d'acceptation atteints avec l'instance (01)4-20 (150 sessions) . . . . .	133
5.27 Instances pour le dimensionnement des réseaux 3G . . . . .	136
5.28 Dimensionnement des réseaux 3G en fonction de la charge de trafic . . . . .	136
II.1 Valeurs des paramètres pour l'application Voix . . . . .	151
II.2 Valeurs des paramètres pour l'application Vidéo conférence . . . . .	151
II.3 Valeurs des paramètres pour l'application Flux vidéo . . . . .	152
II.4 Valeurs des paramètres pour l'application Navigation Internet . . . . .	152
II.5 Valeurs des paramètres pour l'application Courrier électronique . . . . .	152
III.1 Délais pour une session de navigation Internet . . . . .	154
III.2 Délais pour une session de courrier électronique . . . . .	154
III.3 Débits minimum pour le flot $f_s^{\text{DL-WL}}$ d'une session de navigation Internet . . . . .	155
III.4 Débits minimum pour le flot $f_s^{\text{DL-WL}}$ d'une session de courrier électronique . . . . .	155

## LISTE DES FIGURES

1.1	Architecture d'un réseau d'accès CDMA2000 à commutation de paquets . . .	4
2.1	Profil de trafic sur une journée selon Bertini <sup>[1]</sup> . . . . .	26
3.1	Architecture du réseau modélisé . . . . .	33
3.2	Ordonnanceur WFQ . . . . .	34
3.3	Principe du filtre en seau percé . . . . .	36
3.4	Décomposition d'une session en 4 flots : $f^{UL-WL}$ , $f^{DL-WL}$ , $f^{UL-RL}$ et $f^{DL-RL}$ . . . .	52
3.5	Séquencement temporel : les débuts et fins de sessions ne correspondent pas avec les périodes . . . . .	55
3.6	Séquencement temporel : les débuts et fins de sessions correspondent avec les périodes . . . . .	56
3.7	Répartition uniforme par type d'application des sessions commençant à la période h . . . . .	66
3.8	Sélection des stations de base à installer physiquement pour supporter le trafic sur toutes les périodes . . . . .	67
5.1	Sélection des stations de base potentielles pour un usager mobile . . . . .	98
5.2	Exemple de réseau réseau01 . . . . .	109
5.3	Exemple de réseau réseau02 . . . . .	110
5.4	Séquencement temporel des sessions dans l'instance (01)3-16 . . . . .	112
5.5	Positions des stations de base activées dans les instances (01)1-20 et (01)1-30	122
5.6	Taux d'occupation des stations de base sur les périodes les plus chargées .	124
5.7	Répartition des usagers en <i>soft</i> handover . . . . .	128
5.8	Evolution des taux d'occupation des stations de base suivant le profil de trafic	129
5.9	Variations sur la durée de planification du niveau de service atteint pour les applications de Voix . . . . .	134
I.1	Ordre des retransmissions avec le protocole SRP . . . . .	147
I.2	Délai sans retransmission . . . . .	148

I.3	Délai avec une retransmission . . . . .	149
I.4	Délai avec 2 retransmissions . . . . .	149
I.5	Délai avec 3 retransmissions . . . . .	150

## LISTE DES NOTATIONS ET DES SYMBOLES

- 3G : 3rd Generation
- ARIB : Association of Radio Industries and Business
- BS : Station de base
- BSC : Base Station Controller
- CAC : Call Admission Control
- CBR : Constant Bit Rate
- CDM : Code Division Multiplexing
- CDMA : Code Division Multiple Access
- DHCP : Dynamic Host Configuration Protocol
- DS-WCDMA : Direct Sequence Wide band CDMA
- ETSI : European Telecommunications Standards Institute
- FA : Foreign Agent
- FDD : Frequency Division Duplex
- HA : Home Agent
- GoS : Grade of Service
- GSM : Global System for Mobile communication
- IETF : Internet Engineering Task Force
- IP : Internet Protocol
- IPM : IP mobile
- ITU : International Telecommunications Union
- MC-CDMA : Multi Carrier CDMA
- PCF : Packet Control Function
- PDSN : Packet Data Serving Node
- PGPS : Packet Generalized Processor Sharing
- PLM : Problème Linéaire Mixte
- PPP : Point to Point Protocol
- PSTN : Public Switch Telephone Network

- QoS : Qualité de Service
- RAB : Radio Access Bearer
- RAN : Réseau d'Accès Radio
- RC : Radio Configuration
- RCC : Radio Ressources Control
- RL : Radio Link
- RLP : Radio Link Protocol
- RPPS : Rate Proportional Processor Sharing
- RTP : Real Time Transport Protocol
- SIP : Session Initiation Protocol
- SIR : Signal Interference Ratio
- SRP : Selective Repeat Protocol
- TIA : Telecommunications Industry Association
- TCP : Transmission Control Protocol
- TDD : Time Division Duplex
- TDM : Time Division Multiplexing
- TTA : Telecommunications Technology Association
- UDP : User Datagram Protocol
- UMTS : Universal Mobile Telecommunications System
- VBR : Variable Bit Rate
- VoIP : Voix sur IP
- WFQ : Weighted Fair Queuing
- WL : Wired Link

Quand la traduction française n'est pas suffisamment explicite ou que l'anglicisme est largement utilisé dans les documentations et articles francophones, le terme anglais sera conservé (ex: *soft handover*).



## LISTE DES ANNEXES

<b>Annexe I:</b>	<b>Protocole de Répétition Sélective . . . . .</b>	<b>147</b>
I.1	Principe . . . . .	147
I.2	Délais moyens de transmission . . . . .	148
<b>Annexe II:</b>	<b>Paramètres du générateur de trafic . . . . .</b>	<b>151</b>
II.1	Paramètres pour la Voix . . . . .	151
II.2	Paramètres pour la Vidéo conférence . . . . .	151
II.3	Paramètres pour le Flux vidéo . . . . .	152
II.4	Paramètres pour la Navigation Internet . . . . .	152
II.5	Paramètres pour le Courrier électronique . . . . .	152
<b>Annexe III:</b>	<b>Exemple de calcul des différents délais liés au flot descen-</b>	
	<b>dant d'une session VBR . . . . .</b>	<b>153</b>
III.1	Calcul des délais pour les sessions de navigation Internet . . . . .	154
III.2	Calcul des délais pour les sessions de courrier électronique dans le sens des- cendant . . . . .	154
III.3	Calcul des valeurs minimales de débit pour le flot descendant imposées par la contrainte de délai. . . . .	154

## INTRODUCTION

### 0.1 Motivations

L'accès à l'information n'importe où et n'importe quand est la ligne de mire des futurs réseaux de communication mobile troisième génération. Ainsi, le grand défi de cette nouvelle technologie est de pouvoir combiner la mobilité avec l'accès à toutes sortes d'application, en particulier les applications multimédia si prisées par cette nouvelle ère de communication. Ces dernières nécessitent de hauts débits de transfert et ont des exigences de qualité de service, ce que les réseaux mobiles actuels de deuxième génération ne sont pas capables de fournir.

Aujourd'hui, même si les réseaux fixes peuvent fournir des hauts débits, la notion de qualité de service, en revanche, n'est pas encore complètement maîtrisée. En-effet, les applications à succès du début de l'Internet, tel que la navigation ou FTP, exigeaient un transport fiable des données sans contraintes strictes sur les délais. Ainsi, les modes de transport non orienté connexion ainsi que le service *Best Effort* permettent d'optimiser les ressources en s'accommodant de ces critères. En revanche, les nouvelles applications multimédia, comme la vidéo conférence ou le transfert de flux vidéo temps réel, ont des exigences strictes sur le débit, le délai et même la gigue. De nouveaux mécanismes s'imposent alors pour différencier les services et pour fournir la qualité de service exigée. Parmi ces mécanismes, on compte les contrôles d'admission, les protocoles de réservation de ressources et les politiques d'ordonnancement des paquets. Une autre difficulté pour la troisième génération est de maintenir les hauts débits et la qualité de service sur les liens radio qui sont peu fiables pour les transmissions. La technique de multiplexage CDMA, choisie pour les futurs réseaux, permet de diversifier les débits sur les liens radio et d'augmenter la capacité de trafic des cellules.

Les travaux de recherche sur la troisième génération sont très avancés et certains constructeurs testent déjà physiquement leur architecture. Ainsi, les outils de dimensionnement adaptés à cette technologie sont d'actualité. En effet, les nouveaux profils de trafic seront très gloutons en termes de ressources. Compte tenu de la concurrence accrue et du prix élevé des équipements, un outil de dimensionnement sera primordial pour définir la configuration optimale des équipements au coût minimal.

Notre étude se donne donc comme objectif de définir un modèle d'optimisation pour le dimensionnement de réseaux troisième génération et d'utiliser cet outil de planification pour évaluer le dimensionnement optimal d'un réseau test.

## 0.2 Organisation du mémoire

Le chapitre 1 présente le contexte des réseaux troisième génération et s'attarde particulièrement sur les mécanismes clé et les profils de trafic qui influencent le dimensionnement. Le chapitre 2 décrit les principales approches des modèles déjà existants pour le dimensionnement de réseaux fixes et de réseaux mobiles.

La modélisation menée dans notre étude est décrite en détails dans le chapitre 3 qui s'organise en trois parties :

- la modélisation du réseau comprenant la partie fixe (réseau de cœur) et la partie radio (réseau d'accès radio),
- la modélisation du trafic attendu pour la troisième génération. Les différentes applications et les critères de qualité de service intégrés au modèle sont définis dans cette partie,
- la stratégie de dimensionnement.

Le chapitre 4 est consacré à la formulation mathématique du modèle d'optimisation.

Le chapitre 5 décrit, dans un premier temps, l'implémentation du modèle ainsi que les paramètres du trafic. Le logiciel CPLEX-MIP, utilisé pour résoudre le problème d'optimisation, est détaillé succinctement. Toute une section est réservée à l'étude des stratégies de branchement avec CPLEX-MIP. En-effet, nous mettons l'accent sur la modélisation et la validation du modèle plutôt que sur le développement d'un branchement particulier. Ainsi, nous cherchons une stratégie de CPLEX-MIP la plus adéquate pour résoudre nos instances. Après la validation du modèle, nous utilisons notre outil de dimensionnement pour évaluer les ressources nécessaires aux futurs réseaux de communication mobile.

Enfin, une conclusion est apportée quant à l'ensemble des résultats obtenus et les améliorations à apporter au modèle pour le rendre plus performant sont évoquées.

### 0.3 Contributions

Dans le cadre de notre étude, nos développements ont apporté :

- une modélisation du réseau qui intègre les nouveaux concepts de la troisième génération,
- un modèle qui compose avec la différenciation des services grâce à un contrôle d'admission, à une politique d'ordonnancement des paquets et à des formules de capacité radio multi-services,
- un outil de dimensionnement qui englobe à la fois le réseau de cœur et le réseau d'accès radio,
- une évaluation du dimensionnement optimal d'un réseau test en fonction du trafic escompté,
- un moyen de comparer l'efficacité des contrôles d'admission mis en oeuvre sur les ressources nécessaires en prenant comme référence le dimensionnement obtenu avec le contrôle d'admission idéal, car anticipatif, du modèle.

## CHAPITRE 1

### ARCHITECTURE D'UN RÉSEAU D'ACCÈS CDMA2000 ET SERVICES MULTIMÉDIA

#### 1.1 Les réseaux troisième génération (3G)

##### 1.1.1 Historique

Les deux grands objectifs des réseaux multimédia 3G sont, d'une part, de fournir les mêmes services que les réseaux fixes, impliquant ainsi une convergence entre les deux technologies et, d'autre part, de supporter le *global roaming*. Cet objectif, très ambitieux, a pour but de répondre aux nouvelles exigences de mobilité des clients qui souhaitent utiliser les services de leur terminal mobile à l'étranger. Cela implique l'inter-opérabilité des systèmes et une coopération entre les pays pour fournir les services.

Les systèmes 3G vont ainsi remplacer les systèmes 2G déjà implantés depuis le début des années 90. Les réseaux 2G, dont la norme la plus répandue est le GSM, ont introduit les transmissions numériques sur les liaisons radio, accédant ainsi à des débits permettant le transport de la voix et d'applications de courts messages. Actuellement, la génération 2.5 fournit des applications limitées de navigation (WAP et imode) qui permettent l'accès à des serveurs de données. Cependant, les faibles débits et les terminaux peu adaptés en font une génération intermédiaire. Une procédure de standardisation nommée IMT-2000 a été créée par l'ITU pour intégrer les réseaux 2G déjà existants avec la technologie CDMA, avec les réseaux de satellites et avec les nouvelles technologies associées au transport des applications multimédia. Ce travail de standardisation a été mené en parallèle dans plusieurs pays. Ainsi, en juin 1998, les groupes de recherche ETSI (Europe), TTA (Etats-Unis), ARIB (Japon) et TTC (Corée du Sud) soumettent leurs propositions et en décembre 1998 le groupe commun de standardisation 3GPP est créé. Un des objectifs premiers de 3GPP est d'assurer l'inter-opérabilité entre les systèmes 3G de tous les pays. En mars 1999, la technologie du CDMA est choisie comme la technologie de transmission radio pour tous les systèmes 3G. Cependant, plusieurs variantes sont possibles. Ainsi les Etats-Unis, l'Amérique du Sud et la Corée du Sud ont opté pour MC-CDMA (CDMA-2000) alors que l'Europe et

le Japon ont opté pour DS-WCDMA (UMTS). De plus, certains paramètres changent dans l'implantation d'une même technologie. Ainsi le Japon et l'Europe ont des paramètres techniques différents rendant l'inter-opérabilité difficile. 3GPP2 est l'équivalent de 3GPP pour CDMA-2000. Actuellement, UMTS et CDMA-2000 ne sont pas compatibles mais 3GPP et 3GPP2 harmonisent leurs recherches pour résoudre ce problème. Les premières versions de CDMA-2000 ont été déployées en 2000 en Corée du Sud et en 2001 aux Etats-Unis, au Canada et en Amérique du Sud. On prévoit le déploiement de l'UMTS pour 2004. La coexistence des réseaux 2.5G et 3G s'étendra sur une dizaine d'années.

L'évolution de CDMA-2000 se fait progressivement et la version 1x est compatible avec CDMAone qui est le protocole de transmission pour les réseaux 2G en Amérique du Nord. CDMA-2000 1x permet de supporter deux fois plus de voix en termes de capacité et pousse les débits jusqu'à 144 kbps pour les données. CDMA-2000 3x permet la transmission simultanée de la voix et des données avec un débit allant jusqu'à 2 Mbps. Ces versions sont en cours de développement avec de nombreuses avancées. Ainsi la version CDMA-2000 1xEV-DEV intègre la qualité de service, l'authentification et la voix sur IP (VoIP). De plus, elle permet des débits élevés (jusqu'à 2 Mbps) sur une bande de fréquence de 1.25 MHz.

La structure des organismes et les détails des spécifications peuvent être consultés sur les sites web de 3GPP et 3GPP2<sup>[2], [3]</sup>.

### 1.1.2 CDMA2000 versus UMTS

Les deux normes CDMA2000 et UMTS utilisent CDMA comme technique de multiplexage. Succinctement, CDMA permet à chaque utilisateur d'émettre sur un spectre étalé obtenu au moyen d'un code aléatoire personnel. Tous les utilisateurs utilisent ainsi simultanément la même bande de fréquence. Le signal de base est d'abord converti en un signal radio numérique à bande étroite et se voit ensuite alloué un code d'étalement afin de pouvoir être distingué parmi les signaux d'autres utilisateurs. Deux modes de fonctionnement sont possibles :

- mode FDD où deux spectres de fréquence disjoints sont réservés, l'un pour les liaisons du terminal mobile vers la station de base (liaisons montantes) et l'autre pour les liaisons de la station de base vers le terminal mobile (liaisons descendantes),
- mode TDD où un seul spectre de fréquence est utilisé pour les deux sens.

Le problème de la réutilisation des fréquences dans les cellules voisines, qui se posaient avec les technologies TDMA et FDMA, ne se pose plus avec CDMA. En-effet, tous les terminaux mobiles utilisent la même bande de fréquence et ne sont plus en interférences fréquentielles entre eux grâce aux codes d'étalement orthogonaux. On obtient ainsi une forte efficacité spectrale et la capacité des cellules est augmentée. La source principale des interférences n'est plus liée aux corrélations fréquentielles mais plutôt à des problèmes de puissance. Chaque utilisateur doit avoir le même poids en puissance sur le spectre étalé. Un contrôle de puissance efficace devient donc une nécessité (section 1.3.1). Notons que l'orthogonalité des codes n'est pas parfaite et, de ce fait, les interférences entre les usagers mobiles ne sont pas complètement éliminées. D'après [4], la pseudo-orthogonalité est comprise entre 0.4 et 0.9 si on considère qu'une orthogonalité parfaite est 1.

Le tableau 1.1 résume les principales différences techniques entre UMTS et CDMA-2000.

TAB. 1.1 – *Différences techniques entre UMTS et CDMA-2000*

	UMTS	CDMA-2000
Technique de multiplexage	DS-WCDMA	MC-CDMA
Spectre de fréquence	5 MHz	$N \times 1.25$ MHz ( $N=1,3$ )
Le débit <i>chip</i>	3.84 Mcps	$N \times 1.2288$ Mcps ( $N=1,3$ )
Durée d'une trame	10 ms	20 ms
Fréquence pour le contrôle de puissance	1500 Hz pour les deux sens	800 Hz pour les deux sens
Synchronisation des stations de base	asynchrone	synchrone
Canaux pilotes	TDM	CDM

La différence la plus importante est le spectre disponible: une bande de fréquence de 5 MHz pour W-CDMA et une ou trois bandes de fréquence de 1.25 MHz suivant la version de MC-CDMA (CDMA-2000 1x et CDMA-2000 3x). Le partage du spectre pour les canaux pilotes est aussi différent.

Un avantage de W-CDMA est la possibilité d'offrir des débits variables alors que les débits sont fixes avec CDMA-2000. En revanche, la migration de la 2G vers la 3G est plus simple avec CDMA-2000, les équipements 2G de CDMAone étant réutilisés pour CDMA-2000. C'est un sérieux avantage compte tenu du montant des investissements et ce qui explique, du même coup, l'avance dans l'implantation de CDMA-2000.

## 1.2 Architecture d'un réseau d'accès CDMA-2000

Le réseau d'accès permet à un terminal mobile d'accéder aux réseaux téléphoniques commutés et aux services de données fournis par des réseaux publics. Un réseau d'accès se compose d'un réseau d'accès radio, qui gère les liaisons radio et les handovers, et d'un réseau de cœur, qui route les connexions, qui fait l'interface avec les réseaux externes et qui identifie les usagers. Actuellement, les réseaux de cœur supportent le transport par commutation de paquets pour les données et par commutation de circuits pour la voix. La vision la plus répandue prévoit que les réseaux de cœur seront uniquement destinés à la commutation de paquets (IP) avec pour la voix le protocole VoIP. La figure 1.1 illustre l'architecture d'un réseau d'accès CDMA-2000 à commutation de paquets (d'après [5]).

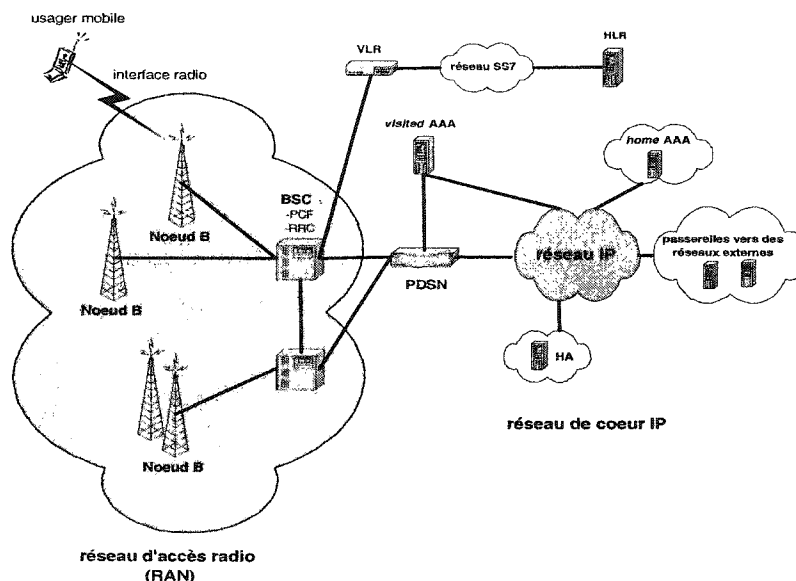


FIG. 1.1 – Architecture d'un réseau d'accès CDMA2000 à commutation de paquets

Les principaux éléments de la figure 1.1 sont :

- **BSC (*Base Station Controller*)**

Le BSC a deux fonctions : le contrôle des paquets (PCF) et le contrôle des ressources radio (RRC). Le PCF établit et maintient les connexions avec le PDSN (*Packet Data Serving Node*). Il communique avec le RCC pour disposer des ressources radio nécessaires au transfert des paquets sur les liaisons radio. Pendant un *soft handover*



(définition à la section 1.3.2) vers une station de base reliée à un BSC différent, le PCF transfère ses informations au PCF qui prendra le relais pour maintenir la communication vers le PDSN. Le RRC communique avec les serveurs d'authentification VLR, HLR et AAA via le PDSN pour attribuer les ressources suivant le profil de l'utilisateur.

– **PDSN (*Packet Data Serving Node*)**

Le PDSN occupe une position centrale. Une des fonctions principales est le routage des paquets vers le réseau de cœur IP ou directement vers le HA (*Home Agent*). Il attribue des adresses IP dynamiques et maintient les connexions PPP jusqu'aux terminaux mobiles. Il regroupe les informations relatives à l'abonné pour l'authentification, les paramètres de la session et les indices de tarification. Il transfère ces informations vers le serveur AAA. En retour, le PDSN récupère le profil de l'utilisateur. De plus, le PDSN joue le rôle de FA (*Foreign Agent*) pour le protocole d'IP mobile (IPM).

– **AAA (*Authentication, Authorization, and Accounting*)**

AAA exécute les requêtes d'authentification envoyées par le PDSN et renvoie en retour une autorisation ainsi que le profil de l'utilisateur mobile. Le PDSN pourra alors gérer la session de façon à respecter les termes du contrat de l'utilisateur (QoS et GoS). Un serveur *home* AAA contient uniquement les données des utilisateurs enregistrés dans ce réseau. Quand un utilisateur se déplace dans un autre réseau d'accès radio, le *visited* AAA (ou Radius) du réseau visité prend le relais. Les spécifications des différents serveurs AAA sont décrites dans la RFC 3141 [6].

– **VLR (*Visited Location Register*) et HLR (*Home Location Register*)**

Les serveurs VLR et HLR contiennent les informations sur l'autorisation et l'authentification des utilisateurs pour le réseau d'accès radio.

### 1.2.1 Le réseau d'accès radio (RAN)

Un nœud B représente une ou plusieurs stations de base. De plus, chaque station de base est composée de trois antennes directionnelles couvrant chacune un angle de 120° : on dit qu'elle est sectorisée. Une cellule est la zone de couverture d'une station de base. Suivant les distances à couvrir et le profil de trafic, plusieurs types de cellules sont définies pour lesquelles les puissances en jeu et les débits disponibles sont différents. Ainsi, les pico

cellules sont utilisées pour les communications à l'intérieur des bâtiments ou dans des zones extérieures confinées, c'est-à-dire des petites zones (dizaines de mètres) à densité élevée avec des usagers à mobilité réduite. Les micro cellules sont destinées aux centres urbains, véritables points chauds avec des utilisateurs nombreux et à mobilité moyenne (10-120 km/h). Les macro cellules couvrent des zones bien plus importantes (zone rurale) avec des rayons de plusieurs kilomètres. La mobilité des usagers est assurée jusqu'à 500 km/h (avions, trains). Pour finir, les cellules globales utilisent les réseaux de satellites pour traverser les océans.

### 1.2.2 Le réseau de cœur

Le réseau de cœur supporte deux protocoles de routage : IP et IP mobile. Avec le protocole IP, le mobile reçoit une adresse IP dynamique attribuée par un DHCP via le PDSN local qui joue le rôle de DHCP intermédiaire pour le mobile. Le mobile retient cette adresse IP tant qu'il se trouve dans un secteur où sa station de base peut communiquer avec ce PDSN. En revanche, si le mobile, de par son déplacement, change de PDSN attribué alors l'adresse IP devra changer. De nombreuses applications comme FTP ne supportent pas le changement d'adresse IP et la connexion sera alors perdue. Pour résoudre ce problème, l'IETF développe un protocole d'IP mobile (IPM) qui permet au terminal mobile de changer de PDSN sans changer d'adresse IP. Une des solutions les plus prometteuses est d'introduire un point fixe dans le réseau qui récupère les paquets envoyés au terminal mobile à son adresse IP fixe (primaire). Ce point fixe est le HA (*Home Agent*). Quand l'utilisateur se déplace vers un réseau radio étranger, le PDSN du nouveau réseau visité joue le rôle d'intermédiaire, nommé FA (*Foreign Agent*), en attribuant une adresse secondaire au terminal mobile. Celui-ci informe alors son HA de cette adresse secondaire et quand ce dernier reçoit des paquets destinés au terminal mobile (adresse IP initiale), il les transfère vers le PDSN étranger (FA) avec l'adresse secondaire en utilisant un tunnel IP. Les paquets sont alors décapsulés par le PDSN puis envoyés vers le terminal mobile. Le changement de réseau est ainsi transparent pour l'entité émettrice. L'avantage de cette solution est qu'elle ne requiert pas de modification des routeurs IP ni des protocoles de transports IP.

### 1.3 Les principaux mécanismes du réseau d'accès radio

Précisons dès maintenant les notations UL et DL qui vont être largement utilisées dans la suite du mémoire. Le sens montant ou UL pour *UpLink* est la liaison radio du terminal mobile vers la station de base. Le sens descendant ou DL pour *DownLink* est la liaison radio de la station de base vers le terminal mobile.

#### 1.3.1 Le contrôle de puissance

Un contrôle de puissance rapide est primordial. Sans lui, un seul terminal mobile émettant à une puissance trop élevée pourrait empêcher tous les autres terminaux mobiles de la cellule de communiquer, puisque plusieurs utilisateurs différents émettent dans la même bande de fréquence. Chaque utilisateur peut être une source d'interférence pour les autres. Il est donc important de mettre en oeuvre un mécanisme qui permette aux terminaux mobiles d'ajuster leur puissance d'émission tout en garantissant une bonne réception à la station de base. Ce problème de puissance se pose aussi pour les puissances émises par la station de base pour limiter les interférences inter-cellulaires. Le contrôle de puissance est donc nécessaire dans les deux sens.

##### **Le contrôle de puissance en boucle fermée**

Ce contrôle de puissance permet de compenser les évanouissements rapides qui dégradent régulièrement le signal. Dans le sens montant, il est basé sur le principe suivant : la station de base réalise des estimations fréquentes du rapport signal sur interférences (SIR) et les compare à une valeur cible. L'expression du SIR est détaillée à la section 3.1.5.2. Si la valeur estimée est supérieure à la valeur cible, la station de base demande au terminal mobile de diminuer sa puissance d'émission. A l'inverse, si elle est inférieure, il sera demandé au terminal d'augmenter sa puissance d'émission. Cette opération est réalisée 1500 fois par seconde (1500 Hz) pour UMTS et 800 fois par seconde (800 Hz) pour CDMA-2000. Ces fréquences élevées permettent au contrôle de puissance de détecter n'importe quelle variation et d'assurer que les puissances émises restent ajustées pour le SIR requis. Ce principe est aussi utilisé dans le sens descendant, bien que, dans ce cas, la raison en soit différente. Dans ce sens, les signaux proviennent de la station de base. Il est souhaitable, afin de minimiser les interférences inter-cellulaires, que la puissance destinée aux terminaux mobiles qui se

trouvent en bordure de cellule soit la plus faible possible tout en garantissant une bonne qualité de réception.

### **Le contrôle de puissance en boucle externe**

Ce contrôle de puissance permet d'ajuster les valeurs cibles des SIR en fonction de l'utilisation du lien radio de façon à assurer une qualité constante. Pour cela, la station de base ajoute aux trames reçues dans le sens montant un indicateur de qualité. Cet indicateur est alors traité par le BSC qui, si la qualité est en baisse, commande en retour à la station de base d'augmenter la valeur des SIR cibles. Cette procédure est implémentée au niveau des BSC car elle doit être toujours disponible même en cas de handover.

### **Le contrôle d'admission en boucle ouverte**

Ce contrôle permet d'évaluer les pertes du canal entre la station de base et l'utilisateur mobile afin de définir à quelle puissance le terminal mobile doit émettre pour compenser les phénomènes d'évanouissements. L'évaluation est faite dans le sens descendant sur des canaux prévus à cet effet et on suppose, de façon abusive, que les pertes sont identiques pour la voie montante et descendante. Bien qu'imprécis, ce contrôle est nécessaire pour les terminaux mobiles lors de l'établissement d'une connexion afin de définir approximativement le niveau de puissance auquel ils doivent émettre. Le contrôle de puissance en boucle fermée permettra ensuite d'ajuster cette puissance.

## **1.3.2 Le handover**

Le handover gère la mobilité des usagers. Notons que les terminaux mobiles et les stations de base utilisent plusieurs récepteurs à corrélation qui forment un récepteur de Rake pour récupérer le maximum d'énergie du signal sur les différents trajets empruntés et éventuellement sur les différentes antennes. Il existe deux types de handover suivant la position de l'utilisateur mobile dans la cellule.

### **Le *softer* handover**

Le *softer* handover se produit quand les stations de base sont sectorisées. Ainsi, quand le terminal mobile se trouve dans une zone de couverture commune à deux secteurs adjacents d'une même station de base, les communications avec la station de base empruntent simultanément deux canaux radio, un pour chaque secteur. Deux codes d'étalement doivent alors être utilisés dans le sens DL afin que le terminal mobile puisse distinguer les deux

signaux issus des deux secteurs et on a donc deux connexions simultanées pour cet usager. Dans le sens UL, les signaux provenant du terminal sont reçus par les deux secteurs de la station de base et routés vers le même récepteur de Rake. Les signaux sont ainsi combinés au niveau de la station de base. On compte généralement 5 à 10 % des terminaux mobiles d'une cellule qui sont en situation de *softer* handover.

### **Le *soft* handover**

Durant un *soft* handover, le terminal mobile se trouve dans la zone de couverture commune à deux stations de base. Les communications entre le terminal mobile et les stations de base utilisent simultanément deux canaux radio, un pour chaque station de base. Du point de vue du terminal mobile, il existe très peu de différences entre le *softer* et le *soft* handover. En revanche, dans le sens UL ces deux handovers diffèrent car, dans le cas du *soft* handover, les signaux reçus par les stations de base sont routés et combinés au niveau du BSC. Cela permet au BSC de sélectionner la meilleure trame reçue. Un usager mobile peut être en situation de *soft* handover avec deux, trois ou quatre stations de base. Si l'utilisateur quitte la zone de couverture commune pour se rapprocher d'une station de base, alors cette dernière le prend en charge. Ainsi, le *soft* handover permet de limiter la perte de connexion quand un usager se déplace vers une autre cellule. On considère que 20 à 40 % des usagers sont en situation de *soft* handover. Il est donc indispensable de prendre en compte les connexions supplémentaires dans une cellule dues au *soft* handover lors du dimensionnement du réseau.

Il existe deux autres types de handover : le *hard* handover inter-fréquences qui permet à un terminal mobile de passer d'un spectre de fréquence à un autre et le *hard* handover inter-systèmes qui permet au terminal mobile de passer d'un système à un autre comme d'un mode FDD à un mode TDD ou pour passer à un système 2G comme le GSM (pendant la période de coexistence des deux systèmes).

## **1.4 La qualité de service dans les réseaux 3G**

Une des caractéristiques les plus connues de la troisième génération est sans doute la possibilité d'offrir aux utilisateurs des débits bien supérieurs à ceux actuellement en vigueur : 384 Kbits/s pour les connexions en mode circuit et jusqu'à 2Mbits/s pour les connexions en mode paquet. Ces hauts débits faciliteront l'introduction de nouveaux services comme les

applications multimédia (vidéo conférence ou téléchargement à flux tendu comme le flux vidéo). La qualité de service et la garantie de service sont deux nouveaux concepts que la 3G doit intégrer pour supporter ces nouvelles applications plus exigeantes. Les sections suivantes décrivent les applications et leurs exigences de qualité de service ainsi que les moyens proposés pour que le réseau puisse satisfaire les requêtes.

#### 1.4.1 Les classes de service

La qualité de service est généralement définie par les critères suivants :

- **Délai** : temps écoulé entre l'envoi d'un paquet par un émetteur et sa réception par le destinataire. Le délai comprend les délais de propagation, de transmission, de traitement et d'attente dans les systèmes intermédiaires.
- **Gigue sur le délai** : variation du délai de deux paquets consécutifs.
- **Bande passante minimum** : taux de transfert minimum pouvant être maintenu entre deux points terminaux.
- **Fiabilité** : taux moyen d'erreurs d'une liaison.

Pour fournir la qualité de service, il faut pouvoir différencier les services suivant leur sensibilité à ces critères. La spécification<sup>[7]</sup> de 3GPP sur la qualité de service propose de classer les applications suivant quatre catégories :

##### **Conversationnelle (*Conversational*)**

Les applications de cette classe impliquent deux utilisateurs humains ou plus qui échangent des informations voix et/ou vidéo en temps réel. Les exigences sur le délai sont strictes, ces derniers doivent être suffisamment faibles pour ne pas dégrader la perception humaine du signal (visuelle et auditive).

##### **Diffusion en flux tendu (*Streaming*)**

Les applications de cette classe impliquent un utilisateur humain et un serveur de données. Le transfert d'information se fait depuis le réseau vers l'utilisateur mobile et la connexion est temps réel et asymétrique. La sensibilité au délai est moins stricte pour cette classe que pour la classe Conversationnelle car il n'y a pas d'interactivité entre les deux entités.

##### **Interactive (*Interactive*)**

Les applications de cette classe impliquent un utilisateur humain et un serveur de données ou d'applications. La connexion est, dans ce cas, basée sur le principe du requête-transfert.

La requête se fait depuis le terminal mobile vers le serveur et le transfert depuis le serveur vers le terminal mobile. L'utilisateur attend certes une réponse rapidement mais les délais restent bien plus importants que pour les classes Conversationnelle et Diffusion. La priorité est mise sur la fiabilité car les données transférées ne doivent pas être altérées. Il est donc possible de traiter ces applications comme non temps réel sans dégrader leur qualité de service.

### Tâche de fond (*Background*)

Les applications de cette classe impliquent un utilisateur, le plus souvent un équipement terminal, qui envoie ou reçoit des données en tâches de fond. L'utilisateur à l'origine de la requête n'attend pas de réponse dans une limite de temps fixée. En revanche, l'intégralité des données est primordiale.

Le tableau 1.2 récapitule, pour les quatre classes de trafic, leurs caractéristiques et leurs exigences de qualité de service.

TAB. 1.2 – Les 4 classes de services définies par 3GPP

Classes de service	Caractéristiques	Exemples	Exigence en QoS
Conversationnelle	<ul style="list-style-type: none"> <li>- transport bidirectionnel entre deux personnes</li> <li>- contraintes de QoS basées sur la perception humaine</li> <li>- conserve la relation temporelle entre les entités du flot</li> <li>- temps réel</li> </ul>	voix sur IP jeux vidéos vidéo conférence	faible délai faible gigue
Diffusion en flux tendu	<ul style="list-style-type: none"> <li>- transport unidirectionnel vers une personne</li> <li>- conserve la relation temporelle entre les entités du flot</li> <li>- temps réel</li> </ul>	flux vidéo flux audio multimédia	faible délai
Interactive	<ul style="list-style-type: none"> <li>- transport bidirectionnel entre un usager et un serveur</li> <li>- conserve la charge utile</li> <li>- non temps réel</li> </ul>	navigation Internet accès bases de données	faible taux d'erreur
Tâche de fond	<ul style="list-style-type: none"> <li>- transport unidirectionnel vers une machine</li> <li>- conserve la charge utile</li> <li>- non temps réel</li> </ul>	courrier électronique transfert de fichiers transfert de données	faible taux d'erreur

### 1.4.2 Transport des applications temps réel

La difficulté pour les futurs réseaux 3G est de pouvoir transporter à la fois des applications non temps réel et des applications temps réel. Pour ces dernières, les flux audio et vidéo sont filtrés à l'émission avec des codeurs, qui, entre autres, imposent des bornes sur les débits des flux. Plusieurs codeurs standards pour l'audio sont recommandés par ITU<sup>[8]</sup>. Parmi ces codeurs, nous retiendrons le standard G.729 CS-ACELP à 8 kbps. L'objectif des codeurs de flux vidéo est de maintenir une bonne qualité tout en réduisant la bande passante requise ainsi que la place mémoire. Parmi les codeurs les plus récents cités dans <sup>[9]</sup>, le standard H.263+ est adapté aux réseaux sans fil.

Le transport des données temps réel est plus complexe car ces dernières ont des exigences strictes sur le délai et la gigue. Actuellement, le protocole pour le transport de la voix et de la vidéo dans les réseaux Internet est RTP (*Real Time Transport Protocol*) qui agit au-dessus de UDP. Comme il est envisagé que les réseaux de cœur de la 3G soient à commutation par paquets, ce protocole est repris par ITU comme protocole de transport associé aux terminaux multimédia H.323. Avant le transport des données, les utilisateurs d'applications temps réel doivent être synchronisés dans le temps mais aussi être d'accord sur la nature de la requête et les exigences de qualité de service. H.323 est un protocole de négociation qui permet d'initialiser les connexions en synchronisant les utilisateurs. Il existe un autre protocole d'initialisation de connexion, SIP (*Session Initiation Protocol*), détaillé dans [5], qui est considéré comme plus puissant, plus simple et plus flexible que H.323. Quand la connexion est initialisée, RTP prend le relai pour le transport des données temps réel. RTP numérote les paquets pour que le receveur puisse reconstruire les séquences, il ajoute aussi un champ pour le temps de *playback*. Ainsi, deux flux utilisant des connexions différentes mais provenant d'une même source (flux vidéo et audio pour la vidéo conférence) pourront être synchronisés sur le temps de *playback*. RTP est indépendant de la nature du média (audio ou vidéo), ainsi des protocoles complémentaires sont développés pour rendre le transport plus spécifique (RFC 1890 pour les flux audio et de vidéo conférence, RFC 2190 pour les flux vidéo H.263, RFC 2198 pour les flux de données audio [6]). RTP, en revanche, ne fournit pas de mécanismes pour la qualité de service.

#### 1.4.3 Le service intégré ou comment fournir la qualité de service

Une fois les critères de qualité de service définis, l'objectif pour les réseaux 3G est de fournir la qualité de service adéquate pour chaque requête. Il faut, pour cela, pouvoir différencier les requêtes et les traiter individuellement. C'est du service intégré. Le groupe de travail Intserv (*INTEgrated SERVices*) a été créé en 1994 pour développer une version améliorée de l'Internet qui supporterait aussi bien du trafic temps réel comme la voix et la vidéo que des transferts de données classiques. Quatre mécanismes permettent de fournir la qualité de service :

- le contrôle d'admission qui vérifie si les ressources sont disponibles au niveau de chaque routeur de la route,



- la réservation des ressources,
- le classificateur de paquets qui, au niveau de chaque routeur, détermine à quel flot appartient le paquet,
- l'ordonnanceur de paquets qui émet ensuite les paquets sur le lien en accord avec les ressources réservées.

#### 1.4.3.1 Le contrôle d'admission ou CAC dans les réseaux fixes

Le CAC est le premier mécanisme appliqué à une requête. Il s'assure qu'en acceptant la session, la qualité de service des sessions déjà en cours ne sera pas dégradée et que les ressources sont suffisantes pour satisfaire les exigences de qualité de service de la nouvelle session. Les politiques de CAC diffèrent selon le type de l'application :

##### **Le contrôle d'admission pour les applications CBR (*Constant Bit Rate*)**

Dans ce cas, le CAC est simple car la source est entièrement définie par son débit. La session ne sera acceptée que si le débit demandé par la source, ajouté au chargement courant d'un lien, est inférieur à la capacité du lien et ce, pour tous les liens de la route. La session pourrait être refusée si la route ne permet pas de garantir son délai maximal toléré. En cas de refus, la session est soit reroutée, soit mise en file d'attente pour une acceptation ultérieure si un lien se libère, soit refusée tout simplement.

##### **Le contrôle d'admission pour les applications VBR (*Variable Bit Rate*)**

Le CAC pour ces applications est plus complexe car ces dernières ont des débits variables avec des périodes de rafales. La difficulté est donc de caractériser les sources VBR pour évaluer les ressources nécessaires à leur qualité de service tout en s'assurant que leur comportement variable ne nuiera pas aux autres applications. Suivant la modélisation de la source, plusieurs stratégies de CAC sont proposées :

##### **– CAC basé sur le débit maximal de rafale**

Ce CAC considère la source VBR comme une source CBR dont le débit serait le débit maximal de rafale. Cette stratégie a l'avantage d'être simple et rapide mais elle conduit à une sous-utilisation des ressources et à des refus abusifs des requêtes par rapport aux ressources disponibles en pratique.

– **CAC basé sur le pire cas**

Il est possible grâce à certaines politiques d'ordonnancement des paquets de garantir des débits minimaux, des taux de pertes maximaux et des délais de pire cas, évalués à partir du pire cas où toutes les sessions VBR enverraient des paquets en rafales en même temps dans le réseau. Les ressources sont alors réservées en fonction du pire cas, ce qui entraîne aussi une sous-utilisation des ressources. Un exemple de délai de pire cas est donné à la section 3.1.2.

– **CAC basé sur des garanties statistiques de qualité de service**

Avec ce CAC, les applications doivent tolérer une faible probabilité que leurs exigences de qualité de service ne soient pas respectées à un instant donné. Le comportement des sources est évalué de façon statistique pour éviter les sur-estimations liées aux CACs précédents. Plusieurs modèles sont proposés dans la littérature pour évaluer la bande passante équivalente d'une source VBR et la capacité équivalente de plusieurs sources VBR multiplexées. La capacité équivalente permettra au CAC d'évaluer la bande passante résiduelle et les ressources à réserver pour la nouvelle requête. Le modèle le plus répandu est celui d'une source markovienne à deux états : état où la source n'émet pas et l'état où elle émet. En supposant que les périodes de rafales et d'inactivité sont distribuées exponentiellement, la source est complètement caractérisée par trois paramètres :

- le débit maximal de rafale,
- la fraction du temps où la source est active,
- la durée moyenne d'une période de rafales.

Le calcul de la bande passante équivalente tient compte du taux de pertes qui est modélisé par le débordement des mémoires tampons allouées. La taille de ces mémoires intervient donc dans l'évaluation de la bande passante équivalente. L'approche statistique rend les formules compliquées. Plusieurs études proposent des approximations simples pour calculer des bandes passantes équivalentes d'une connexion isolée et de plusieurs connexions multiplexées. En-effet, la difficulté est de pouvoir implémenter facilement l'évaluation des ressources nécessaires et disponibles pour faciliter le CAC. Dans cette optique, Guérin *et al.*<sup>[10]</sup> décrivent une métrique simple qui permet au CAC d'opérer les calculs simplement sur tous les liens de la route concernée avec un

minimum de place mémoire nécessaire aux données sur les connexions existantes. Pour plus de détails, les différentes approches pour évaluer les bandes passantes équivalentes sont décrites dans <sup>[11]</sup>.

#### – CAC basé sur des mesures du comportement des sources

Ce CAC effectue des mesures sur le comportement des sources acceptées pour éventuellement renégocier les ressources réservées si la source se comporte différemment de ses spécifications initiales (débit moyen, durée des rafales et débit maximal). Le danger de ce type de CAC est qu'il suppose que les mesures sont un bon indicateur du comportement futur de la source. De plus, ce type de CAC est difficile à implémenter en raison des nombreux calculs mais il permet, d'un autre côté, d'optimiser l'utilisation des ressources. Jamin *et al.*<sup>[12]</sup> proposent un CAC basé sur les mesures pour un réseau paquet à intégration de service. Leur CAC s'applique aux applications qui tolèrent un service prédictif, c'est-à-dire avec des violations de qualité de service occasionnelles. Les classes *Interactive* et *Tâche de fond* supportent ce type de service. La politique d'ordonnancement des paquets choisie est WFQ (section 1.4.3.4). La première partie de l'étude développe des formules pour évaluer le délai d'un nouveau flot compte tenu des ressources effectivement utilisées et des paramètres déclarés par la source. De plus, elle développe des formules pour évaluer l'impact de la nouvelle requête sur les bornes de délais des sessions en cours. La deuxième partie de l'étude se concentre sur le processus de mesure des paramètres qui interviennent dans les formules. La principale difficulté est la durée de la mesure. Il faut trouver un compromis entre des durées courtes et la validité des mesures comme données prédictives. Les résultats permettent de conclure que le service prédictif avec une fenêtre de mesure adéquate permet de fournir une qualité de service pour les délais relativement fiable. Ce type de service aura donc très certainement un rôle à jouer dans les futurs réseaux pour optimiser les ressources.

#### 1.4.3.2 Le contrôle d'admission ou CAC dans les réseaux mobiles

Sur les lien radio dans les réseaux mobiles, les ressources s'évaluent en termes d'interférences et de puissance. Le contrôle d'admission se base ainsi sur des calculs de capacité radio pour déterminer si une station de base peut prendre en charge un terminal mobile.

La section 3.1.5 décrit les formules de capacité radio utilisées dans le modèle pour le CAC. Les exigences de qualité de service sur les liaisons radio portent toujours sur le débit et le taux de pertes mais aussi sur le temps d'initialisation de la connexion et sur le taux de pertes dû au handover. En effet, quand un usager mobile en déplacement est pris en charge en handover par une nouvelle station de base qui est déjà surchargée, il se produit une dégradation des connexions déjà existantes dans la cellule ou la connexion de l'utilisateur peut être perdue. Un contrôle d'admission approprié permet d'éviter une renégociation de la qualité de service à chaque handover. Parmi les nombreuses études de la littérature sur le CAC pour les réseaux mobiles, nous retenons deux grandes approches :

- la méthode de “*Shadow Cluster*”

Cette méthode<sup>[13]</sup> base le CAC sur les ressources disponibles dans la cellule qui prend en charge la requête mais aussi sur l'état de chargement des cellules voisines. Une cellule a un poids d'autant plus grand dans la décision du CAC que l'utilisateur mobile a une forte probabilité de se retrouver dans cette cellule. Les stations de base échangent des informations sur les exigences de qualité de service, sur la position et les prévisions de déplacement. Les stations de base peuvent ainsi évaluer la demande dans un futur proche. Le CAC base alors sa décision sur ce pronostic de futur chargement pour accepter ou refuser une nouvelle requête.

- la partition des canaux

Lee<sup>[14]</sup> propose une partition de la bande passante suivant la nature de la connexion. L'idée est de réserver deux fractions sur la bande passante résiduelle : une fraction pour les nouvelles requêtes et pour les connexions qui seront nouvellement prises en *soft* handover par la station de base et une autre fraction pour les négociations à la hausse des ressources pour une connexion en cours de *soft* handover. La probabilité de perdre une connexion lors d'un handover est ainsi fortement réduite. De plus, Lee suppose que les applications sont flexibles et qu'elles peuvent revoir leurs exigences à la baisse en cas de congestion.

#### 1.4.3.3 La réservation des ressources sur les liens câblés

La réservation des ressources est un mécanisme de signalisation entre les entités émettrice et réceptrice. Interserv a adopté RSVP (*Ressources reSerVation Protocol*) comme

protocole de signalement. Ce protocole est particulièrement adéquat pour les connexions à plusieurs utilisateurs (plus de deux), comme c'est souvent le cas avec la vidéo conférence. Avec ce protocole, l'émetteur envoie un *PATH*-message qui informe les routeurs de la création d'une route. A la réception de ce message, le destinataire renvoie un message de réservation qui contient l'information sur la demande en ressources du destinataire. Avec ce protocole, c'est donc le destinataire qui initie la réservation. Le message de réservation dicte aux routeurs les ressources à réserver sur la route. Si un nouveau destinataire veut se joindre au groupe en faisant une demande de connexion, il envoie un message de réservation jusqu'à un routeur déjà activé par le *PATH*-message. La réservation des ressources se fait alors jusqu'à ce routeur, les ressources étant déjà réservées depuis ce routeur jusqu'à l'entité émettrice. La réservation des ressources est ainsi associée au groupe des utilisateurs plutôt qu'à une connexion.

#### 1.4.3.4 Ordonnancement des paquets

Une fois les critères de qualité de service définis et les ressources réservées, l'objectif est de fournir la qualité de service appropriée pour chaque requête. Il faut, pour cela, pouvoir différencier les requêtes et les traiter individuellement. Une classification des paquets est effectuée au niveau des routeurs pour les placer dans la file d'attente attribuée à la connexion. Puis, l'ordonnanceur de paquets sert chaque file d'attente en accord avec les ressources réservées pour la connexion. Un ordonnanceur de paquets doit pourvoir répondre aux exigences suivantes :

- la simplicité pour faciliter le traitement de chaque file d'attente,
- l'équité et la protection entre les connexions,
- les garanties (déterministes ou statistiques) pour le débit, le délai, la gigue et le taux de pertes,
- l'efficacité dans le partage des ressources pour le contrôle d'admission.

Actuellement, la politique d'ordonnancement des paquets implémentée dans les routeurs des réseaux fixes est la discipline du *Premier Arrivé Premier Servi*, encore appelée FIFO. Cette discipline ne gère qu'une seule file d'attente, commune à toutes les connexions. Elle permet une allocation simple des ressources. En revanche, elle ne différencie pas les connexions et ne les protège pas de certaines qui pourraient être non conformantes, c'est-à-dire trop glou-

tonnes. La garantie de qualité de service est ainsi rendue impossible avec cette discipline. Plusieurs travaux de recherche ont donc été entrepris pour élaborer des politiques d'ordonnement de paquets qui permettraient à la fois de différencier les services et de garantir la qualité de service tout au cours de la connexion.

Le tableau 1.3 résume les avantages et inconvénients des principales politiques d'ordonnement.

TAB. 1.3 – *Avantages et inconvénients des principales politiques d'ordonnement de paquets*

Nom	Conservation de charge	Débit	Délai	Gigue	Avantages	Inconvénients
FIFO	non	non	non	non	- rapide et simple	- pas de protection - ne fournit pas de QoS - peut augmenter la congestion
WFQ	oui	oui	oui	non	- protection - fournit une QoS partielle	- calcul complexe - beaucoup de mémoire pour les données - beaucoup de bande passante pour assurer des délais courts
WF2Q	oui	oui	oui	non	- comme WFQ - plus équitable que WFQ	- comme WFQ - plus complexe que WFQ
JEDD	non	oui	oui	oui	- performant pour la QoS - simplicité des calculs - Taille des mémoires tampon fixe pour une connexion	- réservation au débit maximal de rafales - beaucoup de mémoire pour les données - beaucoup de mémoire tampon - implantation complexe - délais importants pour garantir la gigue
RC	oui ou non	oui	oui	oui	- performant pour la QoS - délai et débit découplés - pas de réservation au débit maximal de rafales - implantation simple - peut fournir les mêmes bornes de délai que WFQ avec des régulateurs supplémentaires	- délais importants pour garantir la gigue - les régulateurs compliquent l'implémentation

Un ordonnanceur à conservation de charge sert les files d'attente tant qu'il reste des paquets dans le système. Ainsi, si une file d'attente est vide, l'ordonnanceur répartit les ressources allouées à cette file sur les autres files pleines. Un ordonnanceur sans conservation de charge reste inactif le temps de service prévu pour cette file et sert les autres ensuite. L'ordonnanceur à conservation de charge est ainsi plus performant pour gérer les applications VBR car il optimise les ressources même pendant les périodes d'inactivité.

Les applications temps réel ont des exigences strictes sur la gigue. Un moyen de la contrôler est de stocker les données au niveau du destinataire, ce qui demande un place mémoire importante. La taille de cette mémoire pourrait alors être réduite si le réseau fournissait certaines garanties sur la gigue. De plus, contrôler la gigue permet aussi de réduire les mémoires tampon au niveau des routeurs. Verna *et al.*<sup>[15]</sup> proposent une solution pour contrôler la gigue. Dans leur étude, l'ordonnanceur de paquets au niveau des routeurs est

accompagné d'un régulateur dont la fonction est de reconstruire et préserver les relations temporelles initiales entre les paquets. Pour cela, l'ordonnanceur estampille chaque paquet d'une valeur qui correspond à la différence de temps entre le moment où le paquet est effectivement servi et le temps où il aurait dû être servi sans gigue. Le régulateur du routeur suivant sur la route se sert alors de cette information pour classer et prioriser les paquets. L'inconvénient de cette stratégie est la complexité de l'implémentation et les temps de traitement par les régulateurs.

Les performances de chaque politique d'ordonnancement des paquets influencent la stratégie du CAC. Ainsi, dans la plupart des articles sur les CAC, les premières parties consistent à étudier les comportements du trafic dans le réseau avec la discipline d'ordonnancement choisie et à évaluer les critères de qualité de service pris en compte par le CAC. Ils décrivent ensuite leur algorithme et le testent sur une simulation de réseau. Zhang<sup>[16]</sup> fait une étude comparative des performances sur le débit, le délai et la gigue de plusieurs politiques d'ordonnancement étudiées dans d'autres articles.

## 1.5 Conclusions

La technique de multiplexage CDMA, le *soft* handover et le contrôle de puissance sont déjà des concepts clé des réseaux 2G. En revanche, le trafic multi-services, la qualité de service et le *Global roaming* sont de nouveaux défis. L'harmonisation des travaux de recherche (3GPP et 3GPP2), la nouvelle architecture basée sur IP, les contrôles d'admission ainsi que les politiques d'ordonnancement des paquets (plus sophistiquées que FIFO) apparaissent comme des solutions prometteuses. Les modèles développés pour le dimensionnement de ces réseaux doivent tenir compte de la nature multi-services du trafic ainsi que des différents mécanismes qui garantissent la qualité de service car ils ont un impact sur les ressources effectivement utilisées. Dans notre étude, l'architecture étudiée modélise tout le réseau d'accès incluant le réseau d'accès radio et le réseau de cœur. Les classes de trafic décrites par le tableau 1.2 sont chacune représentée par une application type. La qualité de service est prise en compte sous forme de garanties sur les débits et les délais, en revanche la gigue n'est pas modélisée. Les ressources réservées pour les sources VBR se basent sur les bandes passantes équivalentes, le contrôle d'admission fournit donc des garanties statistiques pour ces sources. La politique d'ordonnancement des paquets la plus adaptée à notre approche

est WFQ car la gigue n'est pas modélisée et un ordonnanceur à conservation de charge est particulièrement adéquat pour les garanties statistiques.

Le chapitre suivant explore les travaux déjà effectués sur la modélisation des réseaux de communication.



## CHAPITRE 2

### REVUE DE LA LITTÉRATURE SUR LES MODÈLES D'OPTIMISATION DÉJÀ EXISTANTS

Les modèles d'optimisation pour le dimensionnement des réseaux constituent un enjeu important pour les fournisseurs d'équipements et de services qui cherchent à minimiser les coûts pour remporter un contrat. Ces modèles proposent des formulations mathématiques qui modélisent l'utilisation des ressources dans un réseau pour, ensuite, en déduire le dimensionnement optimal, l'objectif étant de minimiser les équipements nécessaires pour satisfaire une certaine demande de trafic. Le dimensionnement des réseaux d'accès radio ainsi que le dimensionnement des réseaux câblés (réseaux publics IP/ATM etc...) ont chacun été largement traités dans la littérature. Ainsi, ce chapitre récapitule les principaux travaux effectués de manière séparée pour ces deux types de réseaux.

#### 2.1 Dimensionnement des réseaux d'accès radio

Les études sur le dimensionnement des réseaux d'accès radio se concentrent principalement sur le nombre et les positions des antennes de transmission (ou stations de base) nécessaires pour couvrir une certaine région. La plupart des outils de dimensionnement se basent sur les surfaces de couverture des antennes potentielles. Cette approche est nommée *Coverage Based Design*. Elle consiste à minimiser le nombre des antennes et à trouver leurs positions tel que le rapport signal sur interférences (SIR) reçu à une station mobile soit suffisamment élevé pour satisfaire la demande. Le SIR reçu à la station mobile dépend de la puissance du signal à l'émission et de l'atténuation du signal : la validité des modèles qui utilisent cette approche dépend donc de la précision de la fonction d'atténuation du signal. Un exemple de *Coverage Based Design*, étudié par Sherali *et al.*<sup>[17]</sup>, consiste à quadriller l'aire géographique avec les positions des usagers mobiles puis à évaluer la fonction d'atténuation du signal entre ces usagers et les antennes potentielles, représentées dans le modèle par des variables continues qui sont leurs coordonnées spatiales. Les contraintes, qui assurent que le SIR reçu par un usager mobile est suffisant, tiennent compte, pour chaque

position potentielle de l'antenne  $(x,y,z)$ , de la distance et de la nature des obstacles à travers la fonction d'atténuation du signal. Des contraintes peuvent être ajoutées pour interdire certains emplacements pour les antennes. Le modèle tolère des cas de connexions refusées en ajoutant dans la fonction objectif une pénalité quand le SIR limite n'est pas atteint pour un usager. Ainsi, la procédure d'optimisation doit choisir entre certains usagers qui n'auront pas un SIR suffisant et l'ajout d'une antenne supplémentaire. Le modèle s'applique au sens descendant. Il serait possible de l'appliquer au sens montant, il faudrait alors trouver un compromis entre les deux dimensionnements obtenus. Après la modélisation, plusieurs algorithmes d'optimisation non linéaire sont proposés pour résoudre le problème.

Tutschku<sup>[18]</sup> propose une version différente du *Coverage Based Design* en transformant le problème de minimisation en un problème de maximisation des surfaces de couverture. Le modèle de Tutschku prend en compte la densité des usagers sur une surface donnée. L'espace est quadrillé avec des surfaces carrées de taille variable pour lesquelles la demande en trafic sur cette surface représente une valeur fixée de requête par minute. La demande en trafic est alors représentée par des points qui sont les centres de ces surfaces. Ces points sont concentrés dans les régions à forte densité de trafic et plus éparpillés dans les zones moins denses. Cette approche se nomme *Demand Based Design*. Dans ce problème de maximisation, le nombre de stations de base à déployer est supposé fixe, la procédure d'optimisation cherche alors à placer ces stations de base de façon à satisfaire le plus de points de trafic compte tenu du SIR. La formulation générique du modèle est la suivante :

$$\text{Maximiser } \sum_{j \in J} a_j y_j \quad (2.1)$$

avec les contraintes :

$$\sum_{i \in N_j} x_i \geq y_j \quad \forall j \in J \quad (2.2)$$

$$\sum_{i \in I} x_i = p \quad \forall j \in J, \quad \forall i \in I \quad (2.3)$$

où :

$I$  ensemble des positions potentielles pour les stations de base,

$J$  ensemble des points de trafic,

$x_i = 1$  si une station de base est sélectionnée à la position  $i$ ,

$= 0$  sinon,

$y_j = 1$  si le point de trafic  $j$  est couvert par une station de base,

$= 0$  sinon,

$N_j = \{i | f_{ji} \leq PL\}$  ensemble des positions  $i$  pour les stations de base tel que

l'atténuation de signal  $f_{ji}$  jusqu'au point de trafic  $j$  soit en dessous de la valeur

limite PL, assurant ainsi une puissance suffisante à la réception,

$p$  le nombre de stations de base à déployer,

$a_j$  l'intensité de la demande au point de trafic  $j$ .

L'objectif consiste à couvrir le maximum de points de trafic. La contrainte (2.2) dit que si un point de trafic est couvert, alors il y a au moins une station de base qui assure le SIR minimum. La contrainte (2.3) force le nombre de stations de base déployées à être exactement égal à  $p$ . Les coefficients  $a_j$  sont utilisés pour mettre des priorités différentes sur les points de trafic favorisant certaines régions comme le centre ville ou un aéroport.

Ces modèles ont le désavantage de ne pas tenir compte de la capacité des stations de base. En effet, la puissance émise par une antenne est limitée. De plus, dépendant de la technique de multiplexage sur les liens radio et de la sectorisation, le nombre d'utilisateurs par station de base est limité. Lee *et al.*<sup>[19]</sup> proposent un modèle d'optimisation qui intègre à la fois l'approche *Coverage Based Design* en se basant sur le SIR pour définir les surfaces de couverture, l'approche *Demand Based Design* en définissant des zones de trafic (TDA) selon la densité des demandes et l'approche *Capacity Based* en ajoutant des contraintes pour empêcher que les capacités maximales des stations de base soient dépassées. Le modèle utilise la capacité maximale des stations de base dans les réseaux 2G : la technique de multiplexage est CDMA, les stations de base sont sectorisées et la nature du trafic se limite à la voix. La formulation

générique du problème est la suivante :

$$\text{Minimiser } \sum_{i=I_1+1}^{I_1+I_2} c_i z_i \quad (2.4)$$

avec les contraintes :

$$z_i = 1 \quad i = 1, \dots, I_1 \quad (2.5)$$

$$y_{ki} \leq z_i \quad k = 1, \dots, N, \quad i = 1, \dots, I_1 + I_2 \quad (2.6)$$

$$\sum_{i=1}^{I_1+I_2} \sum_{k=1}^N d_k y_{ki} \geq \alpha \sum_{k=1}^N d_k \quad (2.7)$$

$$\sum_{k=1}^N d_k y_{ki} \leq M_i z_i \quad i = 1, \dots, I_1 + I_2 \quad (2.8)$$

$$P(k, i) \geq P_{min} * y_{ki} \quad k = 1, \dots, N, \quad i = 1, \dots, I_1 + I_2 \quad (2.9)$$

$$y_{ki}, z_i \in 0, 1 \quad (2.10)$$

où :

$I_1$  nombre de stations de base déjà déployées,

$I_2$  nombre de stations de base potentielles à ajouter éventuellement pour satisfaire la demande croissante de trafic,

$z_i = 1$  si la station de base  $BS_i$  est sélectionnée par la procédure,

$= 0$  sinon,

$y_{ki} = 1$  si la zone de trafic  $TDA_k$  est couverte par la station de base  $BS_i$ ,

$= 0$  sinon,

$N$  nombre de zones TDA,

$\alpha$  taux d'acceptation des requêtes,

$d_k$  demande en trafic de la zone  $TDA_k$ ,

$M_i$  capacité maximale de la station de base  $BS_i$ ,

$c_i$  coût de la station de base  $BS_i$ ,

- $P(k,i)$  puissance reçue par la zone  $TDA_k$  depuis la station de base  $BS_i$ ,  
 $P_{min}$  puissance minimale nécessaire pour assurer la qualité de service.

L'objectif est de minimiser le coût des stations de base nouvellement installées. La contrainte (2.5) force les  $I_1$  stations de base déjà déployées à être sélectionnées par la procédure d'optimisation. Avec la contrainte (2.6), une zone  $TDA_k$  peut être couverte par une station de base déjà existante ou par une nouvelle station de base sélectionnée par la procédure. Grâce à la contrainte (2.7), le modèle tolère un taux de rejet  $(1-\alpha)$ . La contrainte (2.8) impose aux capacités des stations de base une limite maximale. Finalement, la contrainte (2.9) impose que la puissance reçue  $P(k,i)$  par la zone  $TDA_k$  est supérieure à la valeur limite  $P_{min}$ . La puissance reçue dépend, entre autres, de l'atténuation du signal, le modèle considère que l'atténuation est proportionnelle à la puissance quatre de la distance entre l'émetteur et le récepteur. Lee *et al.* proposent un algorithme de recherche tabou pour résoudre ce problème linéaire en nombres entiers. Ils arrivent ainsi à résoudre des cas contenant 2500 TDAs avec une tolérance de 10-20 %.

## 2.2 Dimensionnement des réseaux câblés

Le dimensionnement des réseaux câblés, dans la littérature, consiste à déterminer les capacités des liens de façon à satisfaire la demande de trafic au moindre coût. Généralement, les modèles se concentrent sur la modélisation du taux de rejet, du multi-routage et de la distribution temporelle du trafic (approche *Multi Hours*). Medhi *et al.*<sup>[20]</sup> proposent un modèle de dimensionnement pour les réseaux ATM basé sur la commutation de circuits. Leur stratégie consiste à déterminer les capacités des liens qui minimisent le coût total du réseau avec la possibilité de choisir des chemins virtuels différents pour une classe de service suivant la période de la journée. Le profil de trafic est découpé en plusieurs chargements définis pour chaque classe de service, pour chaque paire Origine-Destination et pour chaque grande période de la journée. Ainsi, pour chacun de ces chargements, la procédure d'optimisation a le choix d'un chemin virtuel parmi un ensemble prédéfini. Une seule route est choisie pour un type de trafic entre une paire Origine-Destination à une période donnée et cette route peut changer pour ce même type de trafic à une période différente. L'approche *Multi Hours*

permet d'optimiser les ressources de façon à satisfaire la demande sur toute la journée. En effet, dimensionner sur la période la plus chargée ne suffit pas car le dimensionnement résultant pourrait ne pas suffir quand la demande varie entre les paires Origine-Destination. L'approche *Multi Hours* devient encore plus importante avec le multi-services. En effet, suivant la période de la journée, le nombre des usagers ainsi que la nature de leurs requêtes changent. Par conséquent, les périodes les plus chargées varient suivant le type d'application. Cependant, les premières estimations supposent actuellement que les périodes les plus chargées pour la voix et pour les autres applications multimédia coïncident<sup>[1]</sup>. Ceci peut se comprendre pour les débuts de la 3G, quand on imagine que quelle que soit l'application, l'utilisation des réseaux sans fil devient généralement nécessaire en dehors des heures de bureau et particulièrement en début de soirée. La figure 2.1 provient de <sup>[1]</sup> et illustre une distribution relative du trafic sur une journée. Le pic du début de soirée est particulièrement marqué, il représente environ 20 % du trafic total quotidien.

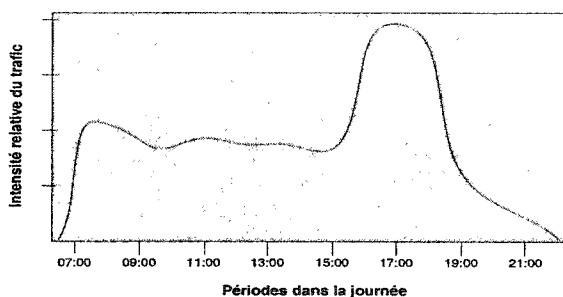


FIG. 2.1 – Profil de trafic sur une journée selon Bertini<sup>[1]</sup>

Medhi et Guptan<sup>[21]</sup> proposent un modèle basé sur la commutation de circuits en considérant du routage dynamique, un trafic multi-services et des taux de rejet spécifiques aux classes de service. Le routage dynamique consiste à proposer un ensemble de chemins de secours entre deux commutateurs en plus du chemin direct. Dans ce modèle, les chemins de secours comportent au plus deux liens. De nouveau, le trafic est découpé en plusieurs chargements (en Erlangs) correspondant à la demande de trafic associée à un type de service, entre une paire Origine-Destination et à une période donnée. L'unité de bande passante est notée BBU (*Basic Bandwidth Unit*) et le nombre de BBU pour une connexion d'un type donné est notée SU (*Service Unit*). Les applications sont acceptées à leur débit maximal de rafale.

La première étape du modèle consiste à évaluer le chargement effectif en unités de service SU pour chacun de ces chargements en considérant un certain taux de rejet (formule inverse d'Erlang-B). La deuxième étape du modèle consiste à passer aux bandes passantes pour évaluer les capacités des liens. Les unités de service effectives sur les liens sont converties en bande passante (BBU) suivant le SU associé à chaque type de trafic. Il aurait été possible d'aggréger le trafic en considérant chaque type d'application directement en unités de BBU et de calculer avec la formule inverse d'Erlang-B le nombre de BBU effectivement nécessaire compte tenu des taux de rejets. Cependant, cette approche entraînerait une sous-estimation des capacités car une des propriétés de cette formule indique qu'une connexion de  $x$  BBU n'est pas équivalente à  $x$  connexions de 1 BBU du point de vue des taux de blocage. Ainsi, le SU est un intermédiaire qui permet d'évaluer les demandes effectives après avoir considéré le taux de blocage dans un contexte de multi-services. La formulation du modèle est la suivante :

$$\text{Minimiser } \sum_{\ell \in \mathcal{L}} c_\ell y_\ell \quad (2.11)$$

avec les contraintes :

$$\sum_{m \in P_k^{sh}} x_{km}^{sh} \geq v(a_k^{sh}, B_k^{sh}) \quad s \in S, \quad k \in K, h \in H \quad (2.12)$$

$$\sum_{s \in S} w_s \sum_{k \in K} \sum_{m \in P_k^{sh}} \delta_{km}^{sh\ell} x_{km}^{sh} \leq y_\ell \quad h \in H, \quad \ell \in L \quad (2.13)$$

$$0 \leq x_{km}^{sh} \leq u_{km}^{sh} \quad m \in P_k^{sh}, s \in S, k \in K, h \in H \quad (2.14)$$

$$y_\ell \leq 0 \quad \ell \in L \quad (2.15)$$

avec :

- $S$  ensemble des différents types de service,
- $K$  ensemble des paires de commutateurs,
- $L$  ensemble des liens câblés,
- $H$  ensemble des différentes périodes de chargement,
- $w_s$  débit maximal de rafale pour un type de service  $s$  (multiples de BBU),

$P_k^{sh}$	ensemble des chemins (chemin direct et chemins de secours) entre une paire de commutateurs $k \in K$ , à la période $h \in H$ ,
$y_\ell$	capacité requise sur le lien $\ell \in L$ (multiples de BBU),
$x_{km}^{sh}$	quantité de chargement en SU effectif pour le chemin $m$ , entre la paire $k$ pour le service $s$ et à la période $h$ ,
$\delta_{km}^{sh\ell}$	paramètre de routage pour indiquer si le chemin $m \in P_k^{sh}$ utilise le lien $\ell$ ,
$a_k^{sh}$	demande en trafic (Erlangs) par le service $s$ à la période $h$ entre la paire $k$ ,
$B_k^{sh}$	Taux de blocage toléré pour le service $s$ , entre les commutateurs de la paire $k$ et à la période $h$ ,
$v(a_k^{sh}, B_k^{sh})$	estimation de la demande en unités de service effectives pour la paire $k$ , le service $s$ à la période $h$ compte tenu du taux de rejet $B_k^{sh}$ ,
$u_{km}^{sh}$	borne maximale sur le chargement du flot associé à la variable $x_{km}^{sh}$ ,
$c_\ell$	coût d'une unité de BBU sur le lien $\ell$ .

L'objectif est de minimiser le coût total des capacités du réseau. La contrainte (2.12) indique que, pour chaque type de trafic  $s$ , entre chaque pair  $k$  et à chaque période  $h$ , plusieurs chemins peuvent être utilisés pour supporter, au moins, la demande effective en unités de service. La contrainte (2.13) permet de passer aux unités de BBU en multipliant les charges effectives (en unités de service), réparties sur les liens, par le SU associé au service et ce pour chaque service et sur chaque lien. La contrainte (2.14) impose une borne maximale sur le flot pris en charge par un lien. L'avantage de ce modèle est qu'il intègre le multi-services avec des taux de rejets spécifiques. Les auteurs valident leur modèle en s'assurant que les taux de blocage réels obtenus dans une simulation d'un réseau (une dizaine de noeuds) dimensionné avec les capacités rendues par leur modèle, sont cohérents avec les taux de blocage  $B_k^{sh}$ .



## 2.3 Conclusions

Les outils actuels de dimensionnement des réseaux d'accès radio se concentrent principalement sur les surfaces de couverture des cellules et les niveaux d'interférence. Bien que ces aspects soient cruciaux pour le dimensionnement, ils ne suffisent pas à garantir des débits spécifiques selon les requêtes. En-effet, les modèles décrits précédemment considèrent uniquement du mono-service (voix), ce qui leur permet d'aggréger les requêtes car elles ont toutes les mêmes exigences de qualité de service (SIR). L'aspect multi-services des réseaux 3G nécessite donc une approche différente. Les outils de dimensionnement pour les réseaux radio 3G devront tenir compte de la multiplicité des applications et des profils d'utilisateurs en fournissant des débits variés. Le facteur limitant est la capacité des stations de base et des formules de capacité radio tenant compte des exigences de qualité de service spécifiques par utilisateur devront être intégrées aux nouveaux modèles.

Certains modèles pour le dimensionnement des réseaux câblés intègrent l'aspect multi-services du trafic en considérant des exigences différentes sur le débit et des taux de rejets spécifiques. En revanche, les critères de qualité de service pour le délai ou la gigue ne sont pas pris en compte. De plus, un des principaux enjeux d'un modèle de dimensionnement global pour le réseau de cœur et le réseau d'accès radio est de coupler ces deux parties. Ainsi, agréger le trafic par type de service comme proposé dans <sup>[21]</sup> en tenant compte des taux de rejet spécifiques ne serait pas exploitable pour les formules de capacité radio qui demandent un traitement par utilisateur. En effet, nous verrons par la suite que la capacité radio requise pour une session dépend, entre autres, de la position de l'utilisateur mobile.

## CHAPITRE 3

### MODÉLISATION ET STRATÉGIE DE DIMENSIONNEMENT

L'objectif du mémoire est la définition, la résolution et l'évaluation d'un modèle d'optimisation pour le dimensionnement d'un réseau 3G. Tout en s'inspirant des concepts de dimensionnement développés dans la littérature (chapitre 2), la modélisation que nous proposons met l'accent sur la différenciation des services. De plus, elle fournit un outil de dimensionnement global qui intègre à la fois le réseau fixe et le réseau d'accès radio.

La pertinence des résultats fournis par la procédure d'optimisation dépend, entre autres, de la granularité de la modélisation. Un modèle trop simple limitera la validité des résultats et un modèle trop complexe rendra la procédure d'optimisation difficile voir impossible. Il ne faut pas aborder la modélisation comme une tentative de reproduction exhaustive de la réalité, elle doit plutôt s'inspirer d'une ligne directrice donnée par les objectifs du modèle. Dans notre étude, on souhaite dimensionner les liens câblés et les stations de base d'un réseau 3G. Il faut donc cibler les éléments principaux de l'architecture, les mécanismes et les protocoles du réseau qui ont une influence sur le dimensionnement et les intégrer le plus simplement possible au modèle.

De plus, la justesse dans la modélisation du trafic est primordiale. Une modélisation abusive entraînerait un sur-dimensionnement ou un sous-dimensionnement. La difficulté vient du fait qu'on ne fait pas de la simulation, il faut prédire le comportement des sources, ce qui est rendu critique avec des applications sporadiques comme la navigation sur Internet ou intermittentes comme la voix.

Enfin, il faut modéliser la stratégie du dimensionnement. Pour les liens câblés, nous avons repris la modélisation couramment utilisée dans les modèles d'optimisation pour le dimensionnement de réseaux câblés comme les réseaux publics (section 2.2). Pour la partie radio, nous proposons une stratégie pour dimensionner en nombre et positions les stations de base.

Le présent chapitre décrit la modélisation du réseau et du trafic ainsi que la stratégie de dimensionnement des stations de base. Une version intermédiaire de la modélisation est développée dans le rapport *Delivrable #2: Definition of an Optimization Model for the Di-*

*mentioning of 3G Multimedia Networks* (février 2002), les grands principes y sont évoqués. Pour certains d'entre eux, le présent mémoire apporte des améliorations. En-outre, certains aspects de la modélisation, comme le dimensionnement des mémoires tampons au niveau des routeurs, ont été conceptualisés sans être approfondis. Ces idées ont été temporairement mises de côté en attente d'une étude ultérieure.

### 3.1 Modélisation du réseau

Le modélisation du réseau respecte l'architecture ainsi que les grands principes d'un réseau 3G comme la différenciation des services, le contrôle d'admission et la garantie de services. Le réseau est à intégration de services et à commutation de paquets.

Le modèle intègre aussi un certain nombre de protocoles des couches physiques qui influencent le dimensionnement. Ces protocoles sont :

- WFQ, ou *attente équitable proportionnelle*, pour la discipline d'ordonnancement des paquets au niveau des routeurs. Cette discipline a l'avantage de garantir un taux de service minimum et une borne sur le temps d'attente pour le paquet. WFQ est détaillée à la section 3.1.2.
- CDMA, ou *accès multiples par affectation de codes*, pour la technique de multiplexage sur les liaisons radio. Cette technique est prise en compte lors du développement des formules de capacités radio. Ces formules sont détaillées à la section 3.1.5.
- SRP, ou *protocole de répétition sélective*, pour la technique de retransmission sur les liens radio en cas d'erreur. Ce protocole est pris en compte lors de l'évaluation des délais de transmission sur les liens radio. Le calcul des délais est détaillé à la section 3.2.5.
- Le réseau de cœur est orienté connexion et le transport par paquets IP est pris en compte dans l'évaluation des délais sur les liens câblés, aussi détaillée à la section 3.2.5.

Les protocoles des couches supérieures à la couche physique ne sont pas modélisés.

Les sections suivantes décrivent en détail les différents aspects de la modélisation du réseau.

### 3.1.1 Architecture

L'architecture du réseau se compose de deux parties distinctes.

- **La partie radio** comprend l'interface radio, c'est-à-dire les liaisons radio entre les terminaux mobiles et les stations de base. Ces dernières correspondent aux nœuds B de l'architecture 3G. On suppose que chaque station de base ne regroupe qu'une seule antenne radio, ce qui n'est pas forcément le cas pour les nœuds B. La partie radio comprend aussi les BSCs (*Base Station Controller*) qui gèrent les ressources des stations de base qui leur sont reliées ainsi que le *soft* handover. La partie radio s'organise en plusieurs sous-réseaux appelés réseaux d'accès radio ou RAN. Un RAN regroupe un BSC et l'ensemble des stations de base qui lui sont reliées.
- **La partie de cœur** ne comprend que des liens câblés et des nœuds. Ces nœuds dans l'architecture 3G sont des nœuds intelligents, leurs fonctions sont décrites à la section 1.2. Le modèle ne fait pas la différence entre ces nœuds, excepté pour le PDSN et les nœuds d'accès externe. Les protocoles liés aux fonctions de ces nœuds ne sont pas modélisés. La partie de cœur assure la liaison entre la partie radio et des réseaux externes comme Internet ou le réseau téléphonique (PSTN).

La figure 3.1 illustre cette architecture. Les directions montante (UL) et descendante (DL) définies à la section 1.3 pour les liaisons radio sont élargies sur la figure au réseau câblé. Ainsi, le sens montant indique la direction d'un terminal mobile vers un nœud d'accès externe et le sens descendant indique la direction d'un nœud d'accès externe vers un terminal mobile.

### 3.1.2 Politique d'ordonnancement des paquets sur les liens câblés

Pour notre étude, nous avons choisi la discipline d'attente équitable proportionnelle, encore appelée WFQ. Cette discipline d'ordonnancement a suscité de nombreux travaux de recherche<sup>[22], [23], [24], [25]</sup>. Une des raisons à sa popularité est sa flexibilité : elle permet de donner des garanties sur les débits et les délais, elle permet de protéger les flots et peut être utilisée pour le contrôle de congestion. Cette discipline est particulièrement adaptée aux réseaux à intégration de services. L'idée initiale est un algorithme d'attente équitable développé par Nagle<sup>[26]</sup>. Demers *et al.*<sup>[22]</sup> ont proposé une version de cet algorithme offrant

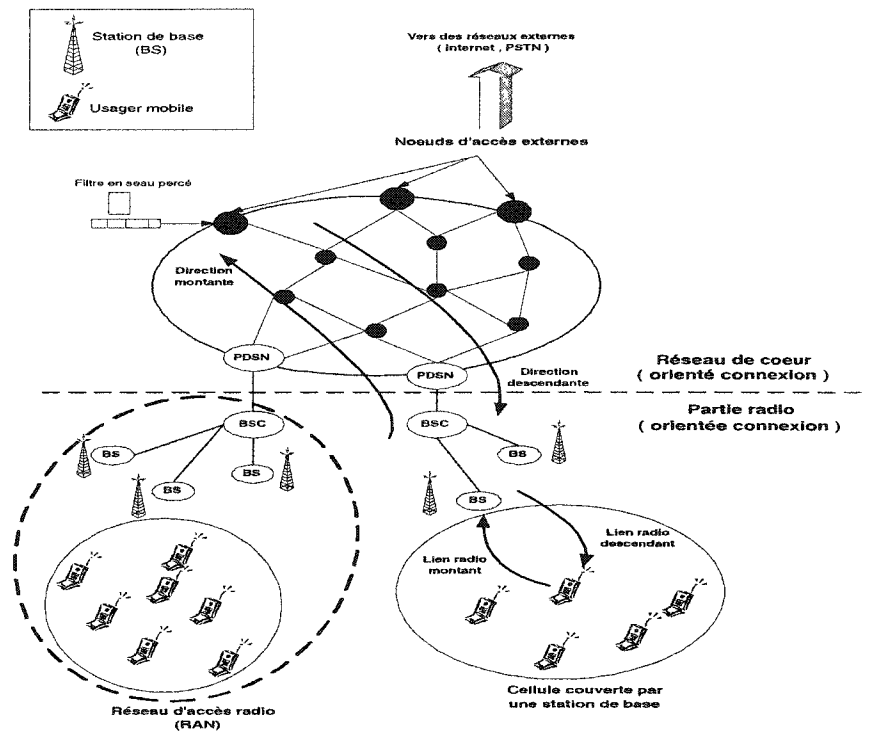


FIG. 3.1 – Architecture du réseau modélisé

une meilleure équité, cette version appelée WFQ a été, par la suite, réinventée par Parekh et Gallager<sup>[23], [24]</sup> sous le nom de PGPS. WFQ et PGPS sont le même algorithme, les concepts énoncés par la suite proviennent de <sup>[23]</sup> et <sup>[24]</sup>, nous utiliserons néanmoins le terme WFQ qui est plus utilisé que PGPS.

Parekh et Gallager étudient les impacts de cet ordonnanceur sur le trafic dans le réseau en considérant le paquet comme une entité indivisible. Dans un premier temps, ils se concentrent sur le cas du nœud simple dans <sup>[23]</sup> puis ils généralisent au cas des nœuds multiples dans <sup>[24]</sup>. Les démonstrations des concepts suivants sont omises mais peuvent être consultées dans <sup>[23]</sup> et <sup>[24]</sup>.

WFQ est une discipline à intégration de services, chaque connexion est classée dans une file d'attente particulière et est affectée d'un coefficient qui détermine son taux de service. La figure 3.2 présente le principe de fonctionnement de l'ordonnanceur WFQ pondéré par les coefficients  $\Phi_1, \Phi_2, \dots, \Phi_N$ . Ces coefficients permettent d'affecter des taux de service différents par file d'attente.

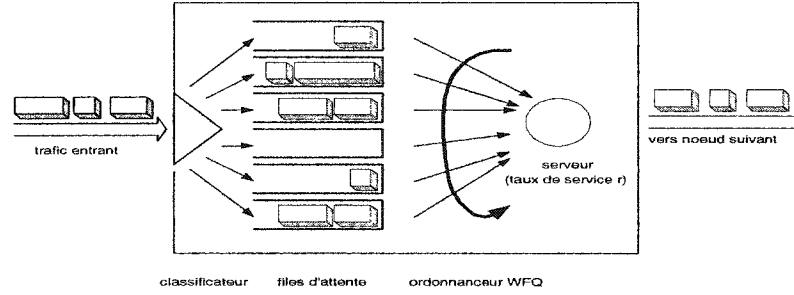


FIG. 3.2 – Ordonnanceur WFQ

Considérons un serveur WFQ avec un taux de service  $r$  constant, servant  $N$  sessions en cours. Le serveur, dans la version paquets de WFQ, traite un certain nombre de paquets sur chaque file qu'il sert à tour de rôle. Une file d'attente est servie proportionnellement à la valeur relative de son coefficient  $\Phi$ . WFQ garantit pour chaque session  $s$  un taux de service minimal :

$$g_s = \frac{\Phi_s}{\sum_{j=1 \dots N} \Phi_j} r. \quad (3.1)$$

WFQ conserve la charge de travail car le serveur est actif tant qu'il reste des paquets dans les files. Si une file est vide, alors le serveur servira la prochaine file et répartira proportionnellement le taux de service affecté à la file vide sur les files pleines. Si toutes les files sont pleines, alors la bande passante allouée à chaque connexion  $s$  est  $g_s$ . Si une file se vide temporairement, alors les autres files seront servies pendant ce temps à un taux plus élevé. Cette propriété de WFQ est particulièrement adéquate et va permettre d'optimiser les ressources. En effet, des applications comme la navigation Internet sont sporadiques et peuvent être très gloutonnes alors que des applications comme la voix sont intermittentes. Les ressources sont protégées contre les sessions gloutonnes et sont redistribuées pendant les périodes d'inactivité.

Parekh et Gallager proposent dans <sup>[24]</sup> une version de WFQ particulière. Les coefficients  $\Phi$ , dans cette version, sont affectés aux sessions en cours de façon à respecter l'ordre relatif des niveaux de qualité de service requis. Cette version se nomme *discipline processeur partagé à taux de service proportionnels* ou RPPS. Avec cette discipline, toute session  $s$  en cours a un coefficient  $\Phi_s$  tel que :  $\Phi_s = \rho_s$  avec  $\rho_s$  le débit déclaré de la session. On obtient le taux

minimal de service garanti suivant :

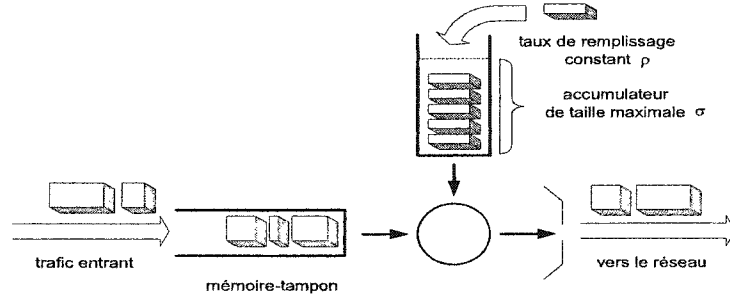
$$g_s = \frac{\Phi_s}{\sum_{j=1\dots N} \Phi_s} r = \frac{r}{\sum_{j=1\dots N} \rho_j} \rho_s. \quad (3.2)$$

Comme le taux d'occupation est forcément inférieur à 1, on a  $\sum_{j=1\dots N} \rho_j \leq r$ . Il vient la propriété suivante :

$$g_s \geq \rho_s. \quad (3.3)$$

Outre la possibilité de servir les sessions proportionnellement à leur niveau de qualité de service, il est possible avec cette discipline de majorer le délai d'attente d'un paquet accumulé à tous les nœuds de sa route. Parekh et Gallager proposent dans [24] une borne sur le délai d'attente de bout en bout. Une des hypothèses pour utiliser cette borne est de contrôler le flot de la session dès son entrée dans le réseau. Ces derniers basent leur étude sur l'utilisation de filtres en seau percé ou leaky buckets pour réguler les sources. De tels filtres, placés à l'entrée du réseau, contrôlent strictement le débit moyen et la sporadicité de la source.

Un filtre en seau percé se compose d'un accumulateur de jetons assimilé à un seau de capacité maximale  $\sigma$  et de taux de remplissage  $\rho$ . Ces deux paramètres définissent complètement les propriétés du filtre. Avec ce principe, les paquets sont admis dans le réseau uniquement après avoir retiré un certain nombre de jetons. En fait, un jeton correspond à une quantité prédéterminée de bits à envoyer dans le réseau. Dépendant de sa taille, un paquet aura besoin de plus ou moins de jetons. Si la taille des paquets est constante, comme la voix par exemple, alors un jeton pourra correspondre au transfert d'un paquet. Le seau est vide quand la session est trop gloutonne et les paquets sont alors mis en attente à l'entrée du réseau jusqu'à ce qu'il y ait à nouveau le nombre suffisant de jetons pour le premier paquet en attente. Le seau se remplit continuellement au taux  $\rho$  tant que la capacité maximale  $\sigma$  n'est pas occupée. Si la source envoie trop de paquets, il est possible que la mémoire tampon qui contient les paquets en attente soit débordée, la source subit alors des pertes. La figure 3.3 illustre le principe du filtre en seau percé caractérisé par les paramètres  $(\sigma, \rho)$ . Ce mécanisme permet de limiter le débit moyen de la source (au plus égal à  $\rho$ ) ainsi que le nombre maximal de paquets envoyés en rafales (au plus égal à  $\sigma$ ). On obtient des

FIG. 3.3 – *Principe du filtre en seau percé*

sources dites conformes.

Pour utiliser la borne sur le délai dans notre modèle, on place un filtre en seau percé pour chaque session à l'entrée du réseau câblé. Pour une session, il y a deux points d'entrée et donc normalement deux filtres :

- pour le flot descendant, le point d'entrée est un nœud d'accès externe,
- pour le flot montant, le point d'entrée est la station de base.

La figure 3.1 ne fait apparaître que le filtre associé au flot descendant. Nous verrons à la section 4.4.4 pourquoi il n'y a pas de filtre pour le flot montant.

Notons  $D_s^*$  la borne maximale sur le délai d'attente d'un paquet d'une session  $s$  qu'il aurait accumulé tout le long de sa route quand la discipline RPPS est appliquée à chaque nœud. D'après [24] il vient :

$$D_s^* = \frac{\sigma_s + (K - 1)L_s}{\rho_s}, \quad (3.4)$$

avec  $(\sigma_s, \rho_s)$  les paramètres du filtre pour la session  $s$ ,  $K$  le nombre de nœuds sur la route et  $L_s$  la taille du paquet.

Cette borne, déterministe, est obtenue en supposant le pire cas, c'est-à-dire quand toutes les sessions sont considérées comme gloutonnes et utilisent le nombre maximum de jetons possible.

### 3.1.3 Contrôle d'admission de bout en bout

Nous proposons, dans notre modèle, un CAC de bout en bout qui base sa décision sur les ressources disponibles à la fois sur les liens radio et sur les liens câblés du réseau. Ainsi,



quand la décision du CAC est positive, les ressources en termes de bande passante sont réservées sur les liens câblés et radio et la discipline d'ordonnancement des paquets RPPS permet de garantir une qualité de service minimum tout au cours de la connexion.

Les ressources nécessaires pour le signalisation ne sont pas prises en compte dans le modèle. De plus, notre CAC n'est pas distribué pour les ressources radio, c'est-à-dire qu'il ne tient compte que des ressources disponibles dans la cellule mère pour accepter la requête sans faire de pronostics sur d'éventuels futurs usagers mobiles en cours de *soft* handover et donc plus prioritaires (section 1.4.3.2).

La réservation des ressources sur les liens câblés se fait au débit déclaré par la source : si la source est de type CBR, ce débit est imposé par la nature du codec et si la source est de type VBR, ce débit correspond à sa bande passante équivalente. Le contrôle d'admission fournit donc des garanties statistiques aux applications VBR. De plus, les débits des flots sur la partie câblée comme sur la partie radio restent constants durant toute la connexion. Les renégociations de qualité de service, effectués par les contrôles d'admission basés sur les mesures, ne sont pas modélisées.

L'algorithme est le suivant :

Considérons une requête pour une session  $s$  initiée par un usager,  $\mathcal{P}$  sa route de l'origine à la destination et  $\rho_s$  sa demande en débit.

– **Test sur les liens câblés**

La bande passante résiduelle  $r$  sur chaque lien câblé de  $\mathcal{P}$  est évaluée au moment de la requête.

Si  $r \geq \rho_s$  pour tous les liens câblés de  $\mathcal{P}$ , alors le test est positif.

– **Test sur le lien radio**

Le niveau d'interférence et la puissance disponible à la station de base sont évalués avec les formules de capacités radio décrites à la section 3.1.5. Le test est positif si la station de base peut prendre en charge la session sans dégrader le niveau d'interférence à un niveau critique (quand l'utilisateur ou la station de base ne peut plus décoder le signal) et sans dépasser la puissance disponible à la station de base. Le test doit être positif pour les liaisons montante et descendante simultanément.

– **Décision d'acceptation ou de refus**

Les deux tests ci-dessus doivent être faits à chaque fois qu'un usager fait une requête.

Cependant, la procédure d'optimisation n'est pas itérative mais globale. Ainsi, elle devra trouver en une seule résolution un scénario où certaines sessions sont acceptées et d'autres refusées. La procédure n'acceptera une session que si les deux tests sont positifs mais la réciproque est fausse. Une session pourra être refusée même si les deux tests sont positifs. En effet, la procédure pourra refuser de telles sessions car elle connaît les requêtes précédentes et suivantes et, si l'une d'elles est plus prioritaire, il pourra refuser une ou plusieurs sessions et réserver ainsi des ressources pour accepter cette session plus prioritaire. Les requêtes avec leur début et leur fin de connexion sont connues car la génération du trafic (les connexions) est déterministe et sert de données d'entrée à la procédure.

La procédure devra dimensionner le réseau de façon à respecter les exigences de qualité de service des sessions acceptées. De plus, le taux de sessions refusées ne devra pas dépasser une certaine limite. On définit ainsi des taux minimaux d'acceptation des sessions. Ces taux sont définis pour chaque paire application-priorité et sont appelés les niveaux de service ou GoS (*Grade of Service*).

Le contrôle d'admission est anticipatif car il peut refuser une session uniquement sur la base d'une session à venir plus prioritaire. Ainsi il tient compte des exigences de qualité de service des sessions mais aussi des priorités pour réserver les ressources et respecter les GoS globaux.

### 3.1.4 Le handover

Nous supposons que les usagers dans la partie radio, c'est-à-dire les terminaux mobiles, ne se déplacent pas. C'est une hypothèse nécessaire pour le routage statique. Le routage est décrit à la section 3.2.3. Le *hard* handover n'est pas modélisé et les stations de base sont supposées non sectorisées. Ainsi, le modèle ne tient pas compte, non plus, du *softer* handover. En revanche, comme suggéré à la section 1.3.2, le *soft* handover est modélisé car il a une incidence importante sur le dimensionnement.

Plus précisément, le *soft* handover a un impact sur le dimensionnement car le terminal mobile occupe de la capacité radio sur plusieurs stations de base. De plus, les puissances émises par le terminal, ou les stations de base, sont plus faibles par rapport aux puissances émises s'il n'y avait pas de *soft* handover. En effet, dans le sens DL, les stations de base envoient au

terminal mobile la même information mais avec des codes d'étalement différents pour que les signaux puissent être distingués. Le terminal récupère alors indépendamment chaque signal puis compare l'information contenue pour obtenir la meilleure qualité possible. Dans le sens UL, le terminal envoie un signal qui est récupéré par les stations de base impliquées dans le *soft* handover, puis les stations de base transfèrent leurs signaux au BSC relié à la station de base mère. Cette dernière, tout comme le terminal mobile, compare les signaux afin de récupérer l'information avec la meilleure qualité possible. Ainsi, dans les deux sens, les transmissions radio peuvent tolérer des taux d'erreur plus importants, la multiplicité des liaisons radio et la comparaison des signaux à la réception aidant à récupérer une bonne qualité. Les puissances émises peuvent alors être plus faibles, ce qui est tout l'intérêt du *soft* handover. En effet, un usager en *soft* handover se situe en général aux extrémités de la zone de couverture de la cellule. Il faudrait donc de fortes puissances pour assurer une bonne qualité sur le lien radio avec la station de base, ce qui dégraderait le niveau d'interférence dans la cellule. Si plusieurs stations de base prennent en charge le mobile, on vient de voir que les liaisons radio établies nécessiteront alors moins de puissance tout en conservant une bonne qualité. Le *soft* handover permet donc de maintenir le niveau d'interférence des cellules à une valeur correcte tout en assurant la connexion d'un usager éloigné.

La puissance à l'émission du signal radio est liée au rapport Energie par bit sur bruit, noté  $E_b/N_t$ , requis à la réception. Un  $E_b/N_t$  trop faible à la réception ne permettra pas à l'entité réceptrice de décoder correctement le signal. Ce rapport est défini plus précisément dans la section suivante consacrée aux formules de capacité radio. D'après les remarques précédentes, le  $E_b/N_t$  requis pour un terminal mobile en *soft* handover est plus faible. D'après *WCDMA for UMTS*<sup>[4]</sup>, le gain du  $E_b/N_t$  requis pour une session en *soft* handover peut varier de 2.7 dB à 4 dB pour un usager mobile à pied (3km/h), selon le facteur d'atténuation entre lui et ses stations de base. Cette réduction est prise en compte dans les formules de capacités radio.

Dans notre modèle, le *soft* handover ne peut se produire qu'entre deux stations de base, pas plus. Les liaisons radio montante et descendante sont assurées en parallèle par les deux stations de base. On suppose que les deux stations de base impliquées sont forcément reliées au même BSC. Cela nous permet d'éviter le multi-routage dans notre modèle, comme nous

pourrons le voir à la section 3.2.3.

De plus, on suppose que les deux stations de base fournissent le même service pour le flot dans une direction. Ainsi, le flot montant ou descendant sera servi avec le même débit et le même taux d'erreur sur les deux stations de base. Par contre, le service dans le sens montant n'est pas forcément le même que dans le sens descendant.

### 3.1.5 Formules de capacité radio

Le service sur un lien radio est caractérisé par deux données :

- une valeur discrète de débit,
- une valeur discrète de taux d'erreur par trame, ou FER.

Ces deux paramètres définissent un RAB (*Radio Access Bearer*). Par la suite, un RAB quelconque sera noté  $r$ , le débit associé  $r_t$  et le FER associé  $r_{\text{fer}}$ . A chaque RAB correspond un rapport Energie par bit sur bruit, ou  $E_b/N_t$ , différent dans le sens montant et descendant. Suivant la technique de modulation, les RAB disponibles sur les liens radio ne sont pas les mêmes. Chaque combinaison de RAB est appelée une configuration radio (RC). Le rapport technique *IS-2000 Forward Link Simulation. Performance Results*<sup>[27]</sup> décrit 9 configurations radio et donne sous forme d'abaques les valeurs des rapports  $E_b/N_t$  associés à chaque RAB. La configuration retenue pour notre modèle est appelée RC4<sup>[27]</sup>.

Le tableau 3.1 donne les valeurs de débits et les valeurs de FER disponibles pour RC4.

TAB. 3.1 – Configuration radio RC4

$r_t$	9.6 kbps	19.2 kbps	38.4 kbps	76.8 kbps	153.6 kbps	307.2 kbps
$r_{\text{fer}}$	0.5%	1%	2%	5%	10%	

Les formules de capacité radio sont développées dans le rapport *Deliverable 2: Call Admission Control Policies and Dimensioning of Multimedia 3G Networks*<sup>[28]</sup>. Ce document s'inspire principalement de *WCDMA for UMTS*<sup>[4]</sup> qui base ses formules sur la technique du WCDMA pour le multiplexage sur les liens radio. Cette technique utilise aussi le CDMA mais sur une bande de fréquence bien plus grande, 5 MHz contre les 1.25 MHz du CDMA2000. Les formules restent donc valides à quelques données numériques près. Les sections suivantes présentent les étapes du raisonnement en omettant les détails de calcul.

### 3.1.5.1 Notations

- $S$ : ensemble des sessions en cours prises en charge par la station de base considérée.
- $\left(\frac{E_b}{N_t}\right)_{s,r}^{UL}$ : rapport de l'énergie par bit du signal reçu à la station de base sur la densité spectrale du bruit reçu (bruit thermique + interférence). Ce signal provient de l'utilisateur mobile associé à la session  $s$ , servi au RAB  $r$  dans le sens montant.
- $\gamma_{s,r}^{UL}$ : valeur cible du rapport  $\left(\frac{E_b}{N_t}\right)_{s,r}^{UL}$ . Cette valeur est la valeur minimale qui permet à la station de base de décoder le signal reçu.
- $\left(\frac{E_b}{N_t}\right)_{s,r}^{DL}$ : rapport de l'énergie par bit du signal reçu par l'utilisateur mobile associé à la session  $s$ , servi au RAB  $r$  dans le sens descendant, sur la densité spectrale du bruit reçu (bruit thermique + interférence).
- $\gamma_{s,r}^{DL}$ : valeur cible du rapport  $\left(\frac{E_b}{N_t}\right)_{s,r}^{DL}$ . Cette valeur est la valeur minimale qui permet au terminal mobile de décoder le signal reçu.
- $SHG^{UL}$ : facteur de réduction de la valeur cible  $\left(\frac{E_b}{N_t}\right)_{s,r}^{UL}$  dans le sens montant quand la session est en *soft* handover.
- $SHG^{DL}$ : facteur de réduction de la valeur cible  $\left(\frac{E_b}{N_t}\right)_{s,r}^{DL}$  dans le sens descendant quand la session est en *soft* handover.
- $\lambda$ : facteur d'interférence. Les interférences inter-cellulaires représentent une fraction  $\lambda$  des interférences intra-cellulaire.
- $LF_{UL}$ : facteur de chargement limite d'un lien radio dans le sens montant.
- $P_T^{UL}$ : puissance totale reçue à la station de base (provenant des terminaux mobiles rattachés à la station de base).
- $P_T^{DL}$ : puissance totale envoyée par la station de base aux terminaux mobiles rattachés.
- $P_s^{UL}$ : puissance émise par le terminal mobile associé à la session  $s$ .
- $P_s^{DL}$ : puissance envoyée par la station de base au terminal mobile associé à la session  $s$ .
- $P_{cont}$ : puissance allouée aux signaux de contrôle et aux signaux pilotes.
- $P_{BS}$ : puissance maximale que peut envoyer une station de base (sens descendant).
- $\nu_s$ : coefficient d'activité de la session  $s$  ( $\nu_s < 1$  pour les applications intermittentes comme la voix).
- $N_o$ : densité spectrale de puissance du bruit thermique.

- $a_{s,0}$ : coefficient d'atténuation du signal entre le terminal mobile  $s$  et la station de base qui la prend en charge.
- $a_{s,i}$ : coefficient d'atténuation du signal entre le terminal mobile  $s$  et une station de base  $BS_i$  voisine.
- $w$ : facteur d'orthogonalité des codes. Les codes utilisés sont supposés être orthogonaux ( $w = 0$ ) pour différencier les usagers. En réalité, les codes ne sont pas parfaitement orthogonaux. Cela a des conséquences sur le niveau des interférences. Holma *et al.* dans *WCDMA for UMTS*<sup>[4]</sup> recommandent de prendre  $w$  compris entre 0.1 et 0.4.
- $Ring(BS_0)$ : ensemble des stations de base voisines de la station de base  $BS_0$  qui prend en charge le terminal mobile.

### 3.1.5.2 Rapport Energie par bit sur bruit pour le trafic multi-services

Le rapport Energie par bit sur bruit doit être suffisamment élevé pour permettre à l'entité émettrice de décoder le signal. L'énergie par bit s'applique au signal utile. Le bruit comprend le bruit thermique et les interférences. Ce rapport est lié à la puissance du signal utile par la formule :

$$\left(\frac{E_b}{N_t}\right)_{s,r} = SIR_s \times \frac{W}{r_t}, \quad (3.5)$$

avec :

- le rapport signal sur bruit  $SIR_s = \frac{\text{Puissance reçue associée à la session } s}{\text{Puissance du bruit et des interférences reçues}},$
- le terme multiplicatif  $\frac{W}{r_t}$ , encore appelé facteur d'étalement.

Dans un contexte de trafic mono-service, les débits sont les mêmes pour toutes les sessions. Le facteur d'étalement est donc identique pour toutes les sessions. Ainsi, les travaux de la littérature sur les capacités radio utilisent indifféremment les valeurs de SIR ou "Energie par bit sur bruit" pour évaluer les capacités radio nécessaires au décodage des signaux. En revanche, pour du trafic multi-services, le facteur d'étalement change avec les sessions. Il est donc nécessaire de se ramener au rapport Energie par bit sur bruit qui sert alors d'unité de mesure commune à toutes les sessions, même si elles sont servies à des débits différents.

### 3.1.5.3 Formules de capacité radio dans le sens montant

Les hypothèses du modèle sont les suivantes :

- les interférences intra-cellulaire et inter-cellulaires sont prises en compte. Les interférences intra-cellulaire représentent les puissances émises par les terminaux mobiles rattachés à la même station de base. Les interférences inter-cellulaires sont évaluées comme un pourcentage des interférences intra-cellulaire<sup>[4]</sup>. L'intensité du trafic doit alors être supposée homogène sur toutes les cellules,
- le contrôle de puissance est supposé parfait,
- les puissances liées aux signaux pilotes et aux signaux de synchronisation ne sont pas prises en compte,
- les cellules ne sont pas sectorisées.

D'après la formule (3.5) et l'expression du SIR décrite à la section précédente, l'expression mathématique du rapport Energie par bit sur bruit appliqué au sens montant s'écrit :

$$\left(\frac{E_b}{N_t}\right)_{s,r}^{\text{UL}} = \frac{W}{r_t} \times \frac{P_s^{\text{UL}}}{((1 + \lambda)P_T^{\text{UL}} - \nu_s P_s^{\text{UL}}) + W N_o},$$

avec :

- $\frac{W}{r_t}$  le facteur d'étalement,
- $P_T^{\text{UL}} = \sum_{j \in S} \nu_j P_j^{\text{UL}}$  la puissance reçue à la station de base,
- $P_T^{\text{UL}} - \nu_s P_s^{\text{UL}}$  les interférences intra-cellulaire provenant des autres sessions en cours,
- $\lambda P_T^{\text{UL}}$  la puissance des interférences inter-cellulaires.

La puissance au numérateur est liée à l'énergie par bit, il faut donc tenir compte de la puissance effectivement reçue quand la session émet une information sans moyenner sur les périodes d'inactivité. En revanche, au dénominateur, les puissances sont liées aux interférences. On tient alors compte des puissances moyennes envoyées par les terminaux mobiles pour évaluer les interférences en multipliant les puissances reçues par les facteurs d'activité associés.

Sachant que le rapport  $\left(\frac{E_b}{N_t}\right)_{s,r}^{\text{UL}}$  doit atteindre la valeur cible  $\gamma_{s,r}^{\text{UL}}$ , on extrait de cette formule une expression de la puissance  $P_s^{\text{UL}}$  requise pour atteindre cette valeur :

$$P_s^{\text{UL}} = \gamma_{s,r}^{\text{UL}} \frac{(1+\lambda) \frac{P_T^{\text{UL}}}{W} + N_o}{\frac{1}{r_t} + \frac{\nu_s \gamma_{s,r}^{\text{UL}}}{W}}.$$

La puissance d'émission du terminal mobile est bornée, cependant on estime qu'il y a une contrainte plus forte liée à la pôle capacité de l'expression qui suit et qui nous permet dévaluer plus justement les capacités requises dans le sens montant. En utilisant  $P_T^{\text{UL}} = \sum_s \nu_s P_s^{\text{UL}}$  et en remplaçant  $P_s^{\text{UL}}$  par la formule ci-dessus, on obtient une expression pour  $P_T^{\text{UL}}$  :

$$P_T^{\text{UL}} = \frac{\sum_s \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}}}{1 - \frac{1+\lambda}{W} \sum_s \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}}} \times N_o.$$

Quand le dénominateur de la fraction ci-dessus devient proche de 0, le système est instable. Pour que  $P_T^{\text{UL}}$  soit défini correctement il faut imposer :

$$\frac{1+\lambda}{W} \sum_s \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}} < 1.$$

Le terme de gauche représente le facteur de chargement  $\eta_{\text{UL}}$  du lien radio.

$$\eta_{\text{UL}} = \frac{1+\lambda}{W} \sum_s \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}}. \quad (3.6)$$

Holma *et al.* suggèrent dans *WCDMA for UMTS*<sup>[4]</sup> de ne pas dépasser un chargement de 60 %. On majore ainsi  $\eta_{\text{UL}}$  par un facteur de chargement limite  $LF_{\text{UL}}$  :

$$\frac{1+\lambda}{W} \sum_{s \in S} \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}} < LF_{\text{UL}}. \quad (3.7)$$

Le modèle utilise la formule (3.7) pour accepter ou refuser une session sur le lien radio montant. Le contrôle d'admission devra ainsi garantir qu'à tout instant la somme des capacités radio des sessions en cours sur une station de base ne dépasse pas le facteur de chargement limite.

Si on considère que la capacité maximale dans le sens montant est  $LF_{\text{UL}} \times \frac{W}{1+\lambda}$ , la contribution de chaque session  $s$  en termes de capacité radio dans le sens montant est :



$$\frac{\gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{UL}} \nu_s}{W}}. \quad (3.8)$$

Si la session est en *soft* handover alors le  $\gamma_{s,r}^{\text{UL}}$  requis est diminué d'un facteur de gain  $SHG^{\text{UL}}$  par rapport au  $\gamma_{s,r}^{\text{UL}}$  sans *soft* handover (section 3.1.4). La capacité radio de la session en *soft* handover devient :

$$\frac{SHG^{\text{UL}} \gamma_{s,r}^{\text{UL}} \nu_s}{\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{s,r}^{\text{UL}} \nu_s}{W}}. \quad (3.9)$$

### 3.1.5.4 Formules de capacité radio dans le sens descendant

On pose les hypothèses suivantes :

- les interférences intra-cellulaire et inter-cellulaires sont prises en compte,
- l'intensité du trafic est supposée homogène sur toutes les cellules,
- le contrôle de puissance est supposé parfait,
- les puissances liées aux signaux de contrôle et de synchronisation sont prises en compte avec la puissance  $P_{\text{cont}}$ ,
- l'atténuation des signaux due aux chemins multiples est prise en compte avec les coefficients  $a_{s,0}$  et  $a_{s,i}$ . L'évaluation de l'atténuation se base sur le modèle d'une double réception du signal :
  - le signal qui a suivi le chemin direct entre l'émetteur et le récepteur,
  - le signal qui a subi une seule réflexion au sol.

Avec ce modèle, la relation entre la puissance envoyée et la puissance reçue est<sup>[29]</sup> :

$$P_r = P_o \left( \frac{h_1 h_2}{d^2} \right)^2, \quad (3.10)$$

avec :

- $P_r$  la puissance reçue,
- $P_o$  la puissance envoyée,
- $h_1$  la hauteur de l'émetteur,
- $h_2$  la hauteur du récepteur,

- $d$  la distance au sol entre l'émetteur et le récepteur.

On tient compte de l'atténuation des signaux dans ce sens et pas dans le sens montant uniquement pour des raisons de définition. L'atténuation est bien un phénomène qui se produit dans les deux sens. Le rapport  $E_b/N_t$  requis est évalué par la station de base pour le sens montant et par le terminal mobile pour le sens descendant. Dans le sens montant, les puissances  $P_T^{UL}$  et  $P_s^{UL}$  sont les puissances reçues au niveau de la station de base, l'atténuation due à la propagation est déjà contenue dans ces termes. En revanche, dans le sens descendant, les puissances  $P_T^{DL}$  et  $P_s^{DL}$  sont les puissances envoyées par la station de base et ce qui compte, c'est ce que reçoit effectivement le terminal mobile, c'est-à-dire les puissances envoyées diminuées par les effets de propagation. D'où les facteurs d'atténuation.

- les cellules modélisées sont hexagonales et seules les 6 premières cellules avoisinantes sont prises en compte dans le calcul des interférences inter-cellulaires.  $Ring(BS_0)$  est l'ensemble de ces 6 stations de base. La puissance des interférences inter-cellulaires n'est plus modélisée par une fraction de la puissance des interférences intra-cellulaire. Les interférences inter-cellulaires deviennent la somme des puissances émises par les stations de base avoisinantes diminuées d'un coefficient d'atténuation. Comme le trafic est supposé d'intensité homogène sur toutes les cellules, on peut supposer que les puissances émises par les stations de base sont toutes égales à  $P_T^{DL}$ .
- Les cellules ne sont pas sectorisées.

D'après la formule (3.5) et l'expression du SIR décrite à la section précédente, et en tenant compte des coefficients d'atténuation, l'expression mathématique du rapport  $E_b/N_t$  appliquée au sens descendant s'écrit :

$$\left(\frac{E_b}{N_t}\right)_{s,r}^{UL} = \frac{W}{r_t} \times \frac{\frac{P_s^{DL}}{a_{s,0}} \times \frac{1}{r_t}}{\left(w \frac{P_T^{DL} - \nu_s P_s^{DL}}{a_{s,0}} + \sum_{i \in Ring(BS_0)} \frac{P_T^{DL}}{a_{s,i}}\right) + WN_0},$$

avec :

- $\frac{P_T^{DL} - \nu_s P_s^{DL}}{a_{s,0}}$  les interférences intra-cellulaire subies par la session  $s$  à la réception au niveau du terminal mobile,

–  $\sum_{i \in \text{Ring}(BS_0)} \frac{P_T^{\text{DL}}}{a_{s,i}}$  les interférences inter-cellulaires provenant des autres stations de base.

On remarque que le facteur d'orthogonalité des codes,  $w$ , pondère les interférences intra-cellulaire. Or ce facteur n'intervient pas dans le cas du sens montant alors que le principe des codes orthogonaux est utilisé dans les deux directions. Notons, dans un premier temps, que si les codes étaient parfaitement orthogonaux, les interférences intra-cellulaire (dues aux autres terminaux mobiles de la cellule) seraient inexistantes. Dans le sens montant, l'orthogonalité des codes, qui permet aux stations de base de différencier les terminaux mobiles, est fortement dégradée par le fait que les émetteurs sont les terminaux mobiles et qu'ils émettent de façon asynchrone les uns par rapport aux autres. Cette pseudo-orthogonalité est responsable des interférences intra-cellulaire. On considère alors le pire cas en supposant que les codes ne sont pas orthogonaux ( $w = 1$ ) et que l'influence des autres terminaux mobiles sur le niveau d'interférence intra-cellulaire est totale. En revanche, dans le sens descendant, les interférences intra-cellulaire sont bien moins importantes car, dans ce cas, l'orthogonalité des codes reste satisfaisante. En effet, la station de base est maintenant l'entité émettrice. Or, cette dernière gère les signaux à l'émission pour conserver leur orthogonalité. Malgré cela, l'orthogonalité n'est tout de même pas parfaite à la réception (par le terminal mobile) car les trajets multiples dégradent les signaux. Les interférences intra-cellulaire sont donc fortement réduites mais ne peuvent être négligées, d'où le facteur d'orthogonalité qui diminue les interférences intra-cellulaire dans le cas du sens descendant. A partir de la formule précédente donnant l'expression du rapport  $\left(\frac{E_b}{N_t}\right)_{s,r}^{\text{UL}}$ , et en considérant que la valeur cible pour ce rapport est  $\gamma_{s,r}^{\text{DL}}$ , on isole  $P_s^{\text{DL}}$  pour obtenir l'expression suivante :

$$P_s^{\text{DL}} = \gamma_{s,r}^{\text{DL}} \times \frac{P_T^{\text{DL}} \left( w + \sum_{i \in \text{Ring}(BS_0)} \frac{a_{s,0}}{a_{s,i}} \right) + W N_o a_{s,0}}{\frac{W}{r_t} + w \nu_s \gamma_{s,r}^{\text{DL}}}. \quad (3.11)$$

La puissance des signaux de contrôle et de synchronisation est prise en compte dans  $P_{\text{cont}}$ .

La puissance envoyée par la station de base est alors :

$$P_T^{\text{DL}} = \sum_{s \in S} \nu_s P_s^{\text{DL}} + P_{\text{cont}}.$$

En remplaçant dans cette expression la valeur de  $P_s^{\text{DL}}$  donnée par l'équation (3.11), on peut extraire une formule pour  $P_T^{\text{DL}}$  :

$$P_T^{\text{DL}} = \frac{P_{\text{cont}} + W N_o \sum_{s \in S} \frac{\nu_s \gamma_{s,r}^{\text{DL}} a_{s,0}}{\frac{W}{r_t} + w \nu_s \gamma_{s,r}^{\text{DL}}}}{1 - \sum_{s \in S} \frac{\left( w + \sum_{i \in \text{Ring}(BS_0)} \frac{a_{s,0}}{a_{s,i}} \right) \nu_s \gamma_{s,r}^{\text{DL}}}{\frac{W}{r_t} + w \nu_s \gamma_{s,r}^{\text{DL}}}}. \quad (3.12)$$

Dans le sens descendant, la puissance  $P_T^{\text{DL}}$  est majorée par une valeur limite  $P_{BS}$  :

$$P_T^{\text{DL}} \leq P_{BS}.$$

En appliquant cette contrainte à l'expression (3.12), on arrive à la formule suivante :

$$\sum_{s \in S} \frac{\gamma_{s,r}^{\text{DL}} \nu_s \left( N_o a_{s,0} + \frac{P_{BS}}{W} \left( \sum_{i \in \text{Ring}(BS_0)} \frac{a_{s,0}}{a_{s,i}} + w \right) \right)}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{DL}} \nu_s}{W}} \leq P_{BS} - P_{\text{cont}}. \quad (3.13)$$

Le modèle utilise la formule (3.13) pour refuser ou accepter une session sur le lien radio descendant. Le contrôle d'admission devra ainsi garantir que cette contrainte ne soit pas violée en acceptant une nouvelle session.

Si on appelle  $P_{BS} - P_{\text{cont}}$  la capacité maximale de la station de base dans le sens descendant, alors la contribution de chaque session  $s$  en termes de capacité radio dans le sens descendant est :

$$\frac{\gamma_{s,r}^{\text{DL}} \nu_s \left( N_o a_{s,0} + \frac{P_{BS}}{W} \left( \sum_{i \in \text{Ring}(BS_0)} \frac{a_{s,0}}{a_{s,i}} + w \right) \right)}{\frac{1}{r_t} + \frac{\gamma_{s,r}^{\text{DL}} \nu_s}{W}}. \quad (3.14)$$

De la même façon que dans le sens montant,  $\gamma_{s,r}^{\text{DL}}$  est diminué d'un facteur de gain  $SHG^{\text{DL}}$  en cas de *soft* handover. La capacité radio de la session en *soft* handover devient :

$$\frac{SHG^{\text{DL}} \gamma_{s,r}^{\text{DL}} \nu_s \left( N_o a_{s,0} + \frac{P_{BS}}{W} \left( \sum_{i \in \text{Ring}(BS_0)} \frac{a_{s,0}}{a_{s,i}} + w \right) \right)}{\frac{1}{r_t} + \frac{SHG^{\text{DL}} \gamma_{s,r}^{\text{DL}} \nu_s}{W}}. \quad (3.15)$$

### 3.2 Modélisation du trafic

#### 3.2.1 Les applications

Dans le contexte des réseaux 3G, il est nécessaire de modéliser les nouvelles applications associées aux réseaux 3G. Pour représenter au mieux la diversité des applications du tableau 1.2, on a choisi dans chaque grande classe de trafic une ou des applications qui deviendront vraisemblablement les plus populaires. Le tableau 3.2 indique les applications que l'on modélise pour la génération du trafic.

TAB. 3.2 – *Les applications du modèle*

Catégorie	Classe de trafic	Application	Flots	Débit
Applications temps-réel	Conversationnelle	Voix	Interactive flots UL et DL, symétriques	CBR
		Vidéo conférence High Vidéo conférence Low	Interactive flots UL et DL, symétriques	CBR
	Diffusion	Flux vidéo High Flux vidéo Low	Non interactive flot DL	CBR
Applications non temps-réel	Interactive	Navigation Internet High Navigation Internet Low	Interactive flots UL et DL, asymétriques	VBR
	Tâche de fond	Courrier électronique	Non interactive flots UL ou DL, symétriques	VBR

Les qualificatifs High et Low portent sur les exigences de qualité de service. Les vidéo-conférences High et Low, par exemple, sont la même application sauf que les exigences en débit et délai seront plus élevées pour l'application High. Prenons les exemples plus particuliers d'un usager classique qui souhaite faire une video conférence avec un ami sur son téléphone et d'un chargé d'affaire qui participe à une réunion à distance sur son ordinateur portable. Les supports ne sont pas les mêmes et la qualité demandée non plus. On différencie ainsi cette même application avec les qualificatifs High et Low.

En suivant le même raisonnement, dépendant des moyens d'un usager, ce dernier souscrit un abonnement lui donnant une priorité plus ou moins grande sur la qualité de service. La qualité de service minimum sera toujours garantie quelle que soit la priorité, mais un usager de forte priorité recevra en moyenne une meilleure qualité. La priorité de l'utilisateur influence aussi la probabilité que sa requête soit refusée par le réseau. On définit ainsi des taux de refus (GoS) tolérés par paires application-priorité. Une forte priorité aura moins de chance d'être refusée. Le profil des usagers ne peut être pris en compte que dans un réseau à intégration de services où chaque requête est traitée individuellement (section 1.4.3). Dans

le modèle, un usager peut être soit de priorité Gold, la plus élevée, soit de priorité Silver. Les applications interactives ont un flot dans les deux sens car l'origine et la destination échangent simultanément des informations. Les applications non interactives n'ont, elles, qu'un seul flot. Le flux vidéo n'a qu'un flot dans le sens descendant car le transfert de données se fait depuis un serveur vers l'utilisateur mobile. Dans le cas du courrier électronique, il n'y a aussi qu'un seul flot mais il peut être dans les deux sens car l'utilisateur mobile peut envoyer ou recevoir un courrier.

Les flots symétriques signifient que les flots dans les sens montant et descendant ont les mêmes exigences de débit et de délai. La voix, par exemple, a les mêmes exigences dans les deux sens. En revanche, une session de navigation Internet est asymétrique car le flot montant, qui correspond à la requête d'ouverture d'une page depuis un serveur, est plus faible que le flot descendant qui correspond au téléchargement de la page. Pour ce flot, le débit sera plus élevé.

On considère dans notre modèle que les applications temps réel ont des débits constants et prédéterminés par le codeur utilisé (section 1.4.2). Elles sont de type CBR. D'un autre côté, les applications qui n'utilisent pas de codeur ont des débits variables. Seules les bornes maximale et minimale sont fixées par le type de l'application. Elles sont de type VBR.

### 3.2.2 Modélisation d'une session

On suppose que les connexions sont toujours entre un usager mobile et un usager fixe appartenant à un réseau externe. D'après l'architecture décrite à la figure 3.1, les sessions sont donc toujours entre un nœud d'accès externe et un terminal mobile. De plus, on suppose qu'un usager mobile ne génère pas plusieurs sessions simultanément.

Par la suite, le réseau sera divisé en deux parties :

- la partie câblée, notée WL, qui comprend tous les liens câblés des nœuds d'accès externes jusqu'aux stations de base,
- la partie radio, notée RL, qui comprend les liens radio entre les stations de base et les terminaux mobiles.

Ces deux parties ne tiennent pas compte de la séparation du réseau global en un réseau de cœur et les RANs. Cette séparation, illustrée à la figure 3.1, se situe au niveau des BSCs alors que la partie câblée va jusqu'aux stations de base.

### 3.2.2.1 Décomposition en flots.

Quand l'application est interactive, la session est modélisée avec deux flots :

- un flot dans la direction montante, noté  $f^{UL}$ ,
- un flot dans la direction descendante, noté  $f^{DL}$ .

Quand l'application n'est pas interactive, la session est modélisée avec un flot dont la direction est celle de la connexion ( $f^{UL}$  ou  $f^{DL}$ ).

Les flots montant et descendant sont eux même divisés en deux flots :

- le flot montant (descendant) sur les liens câblés du réseau, noté  $f^{UL-WL}$  ( $f^{DL-WL}$ ),
- le flot montant (descendant) sur le lien radio, noté  $f^{UL-RL}$  ( $f^{DL-RL}$ ).

La figure 3.4 illustre la décomposition de la session en flots. Dans le cas d'un terminal mobile en *soft* handover entre deux stations de base, les flots  $f^{UL-WL}$  et  $f^{UL-RL}$  sont assurés en parallèle par les deux stations de base.

D'après le contrôle d'admission décrit à la section 3.1.3, la session ne pourra être acceptée par le réseau que si ce dernier a suffisamment de ressources pour accepter les quatre flots simultanément (ou les deux flots si l'application n'est pas interactive). Dans le cas du *soft* handover, il faudra qu'il y ait suffisamment de ressources sur les liens radio établis avec les deux stations de base.

### 3.2.2.2 Attribution d'un débit pour chaque flot

La difficulté dans la modélisation du trafic est de tenir compte de la sporadicité des applications VBR. Or, la procédure d'optimisation va affecter aux flots un débit constant compris entre des bornes maximale et minimale. Les modèles de trafic contournent cette difficulté en calculant une bande passante équivalente compte tenu des caractéristiques comportementales de l'application comme le débit moyen, le débit crête ou la durée des rafales (section 1.4.3.1). Ainsi les sources VBR sont modélisées comme des sources CBR. Les débits fournis par la procédure d'optimisation pour les applications VBR seront donc compris comme des bandes passantes équivalentes.

Dans notre modèle, l'attribution des débits change suivant si l'application est CBR ou VBR.

#### Applications CBR

Les débits sur les liens câblés et sur les liens radio sont fixés par le type de codeur utilisé. La

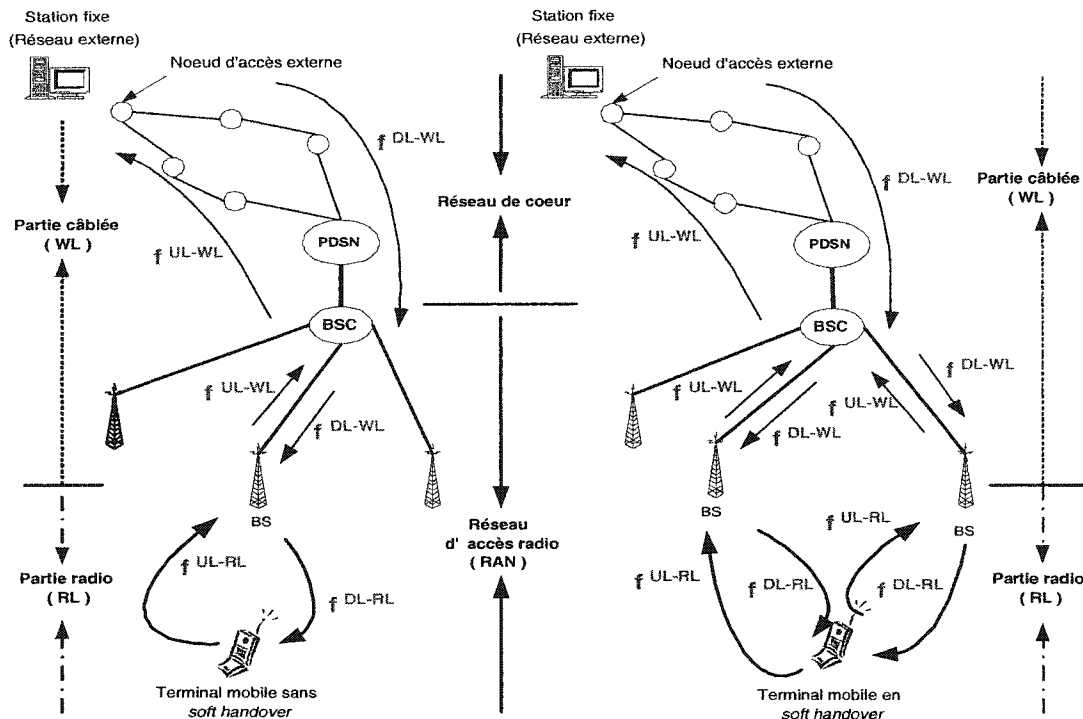


FIG. 3.4 – Décomposition d'une session en 4 flots :  $f^{UL-WL}$ ,  $f^{DL-WL}$ ,  $f^{UL-RL}$  et  $f^{DL-RL}$

procédure d'optimisation devra donc accepter ces sessions aux débits déclarés. Elle aura en revanche le choix entre plusieurs RABs au même débit que celui déclaré mais avec des FERs plus ou moins bons. Ce choix se fera suivant la priorité de l'utilisateur. En effet, dépendant de sa priorité, un usager peut accéder à une certaine qualité de service. Le FER du RAB sélectionné agit sur la qualité de service en influençant le délai de transmission sur les liens radio. Cet aspect est expliqué à la section 3.2.5.

### Applications VBR

Les débits sur les liens câblés et sur les liens radio sont variables. En revanche, des bornes maximale et minimale sont fixées pour chaque flot. La procédure d'optimisation devra donc choisir une valeur continue entre ces bornes pour les liens câblés et un RAB qui respectent ces bornes pour les liens radio.

En plus des bornes maximales et minimales, on impose certaines relations entre les débits des flots d'une session VBR. Ces relations sont les suivantes.

On suppose que dans le sens montant le débit du flot sur le lien radio  $f^{UL-RL}$  est inférieur ou égal au débit du flot sur les liens câblés  $f^{UL-WL}$ . On empêche ainsi une accumulation des



paquets au niveau de la station de base, évitant un délai d'attente non contrôlé pour les paquets.

Dans le cas du sens descendant, on empêche une accumulation incontrôlée des paquets à la station de base en forçant le débit du flot sur le lien radio  $f^{\text{DL-RL}}$  à être supérieur ou égal au débit du flot sur les liens câblés  $f^{\text{DL-WL}}$ .

On impose aussi, par défaut, une valeur minimale  $\mathcal{R}(r_t)$ , pour le flot  $f^{\text{DL-WL}}$  si le RAB  $r$  a été sélectionné pour le flot  $f^{\text{DL-RL}}$ . En effet, sans considération sur le délai, si le flot descendant  $f^{\text{DL-RL}}$  est servi au RAB  $r$  alors le débit sur les liens câblés doit être au moins supérieur au débit du RAB juste inférieur au RAB  $r$  sinon le flot  $f^{\text{DL-RL}}$  sera servi à ce RAB juste inférieur. Par exemple, si le RAB [76.8 kbps,  $x\%$ (FER)] est sélectionné, le débit sur les liens câblés dans le sens descendant devra être supérieur à 38.4 kbps sinon le RAB [38.4 kbps,  $x\%$ ] sera plus adéquat.

Le tableau 3.3 donnent les valeurs  $\mathcal{R}(r_t)$  définies pour chaque débit de RAB  $r_t$ .

TAB. 3.3 – Valeur de  $\mathcal{R}(r_t)$  en fonction du débit sélectionné sur le lien radio descendant

Rab throughput	$\mathcal{R}(r_t)$
9.6 kbps	-
19.2 kbps	10 kbps
38.4 kbps	20 kbps
76.8 kbps	40 kbps
153.6 kbps	80 kbps
307.2 kbps	160 kbps

Les équations suivantes résument les relations entre les débits des flots d'une session VBR:

$$\text{Debit}[f^{\text{UL-WL}}] \geq \text{Debit}[f^{\text{UL-RL}}], \quad (3.16)$$

$$\text{Debit}[f^{\text{DL-WL}}] \leq \text{Debit}[f^{\text{DL-RL}}], \quad (3.17)$$

$$\text{Debit}[f^{\text{DL-WL}}] \geq \mathcal{R}(r_t) \quad \text{quand } \text{Debit}[f^{\text{DL-RL}}] = r_t. \quad (3.18)$$

### 3.2.3 Routage

Une route contient la liste des liens câblés du nœud d'accès externe jusqu'au BSC. Le lien câblé entre le BSC et la station de base qui prend en charge la session n'est pas inclus dans la route car il n'est pas connu au départ. C'est la procédure d'optimisation qui va sélectionner une ou deux stations de base pour la session. Le lien câblé entre le BSC et la station de base dépend donc du choix de la procédure. Le BSC, lui, est connu dès le départ. En effet, le BSC couvre une certaine surface avec l'ensemble de ses stations de base. Si l'utilisateur mobile se trouve sur cette surface alors sa connexion passera par ce BSC.

Les flots montant et descendant d'une session interactive ne prennent pas forcément la même route. Ces deux routes doivent cependant être reliées au même BSC et au même nœud d'accès externe.

Les routes d'une session sont prédéterminées et comme les usagers mobiles sont fixes, la route d'un flot est la même durant toute la connexion. C'est du mono-routage statique.

### 3.2.4 Durée de la planification et séquençage temporel

La durée de planification sur laquelle sont réparties les sessions est fixée, elle devrait être suffisamment étendue pour représenter les différents profils de trafic observés sur une journée.

Comme évoqué à la section 2.2, dimensionner sur la période la plus chargée entraînerait un sur-dimensionnement pour tout le reste de la journée. Le profil d'arrivée des sessions est connu sur toute la durée de la planification. L'approche *Multi Hours* consiste, dans notre modèle, à tenir compte de toutes les sessions de façon globale. Le réseau est dimensionné en fonction de tous les profils observés sur une journée grâce au contrôle d'admission anticipatif. Avec les niveaux de service globaux, les sessions refusées sont réparties au mieux dans le temps pour optimiser les ressources sur toutes les périodes.

Les arrivées et les durées des sessions sont donc des données d'entrée du modèle. Le temps, lui, est divisé en périodes de durée constante. La figure 3.5 illustre un exemple de répartition temporelle des sessions.

Les contraintes de contrôle d'admission (tests proposés à la section 3.1.3) doivent être appliquées à chaque requête. Avec des temps d'arrivée continus, ces contraintes sont alors

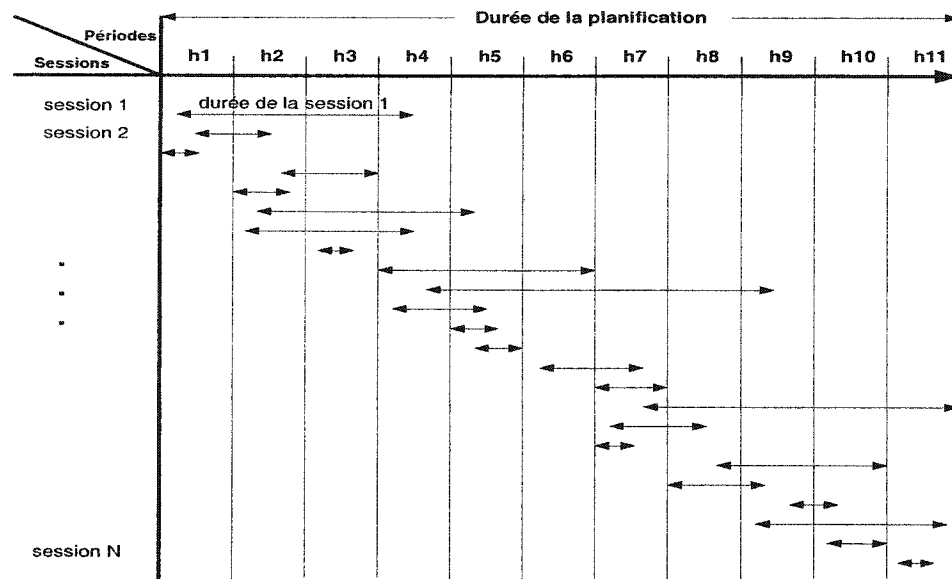


FIG. 3.5 – *Séquencement temporel : les débuts et fins de sessions ne correspondent pas avec les périodes*

multipliées par le nombre de sessions. Pour réduire le nombre de contraintes du modèle mathématique, on transforme les valeurs continues des temps d'arrivée et des durées des sessions en valeurs discrètes. Ainsi, on fait correspondre le début d'une session au début d'une période et sa durée en un multiple de périodes. La figure 3.5 devient la figure 3.6.

Les contraintes de contrôle d'admission s'appliquent ainsi à chaque période et testent l'ensemble des sessions qui font leur demande à cette période. L'état du réseau est donc testé au début de chaque période et après la décision du contrôle d'admission, le nouvel état du réseau est supposé constant jusqu'au début de la période suivante.

Le séquencement temporel est exprimé, dans le modèle, à travers des paramètres de routage qui indiquent, pour chaque session, quels liens câblés du réseau de cœur elle utilise et à quelles périodes.

### 3.2.5 Délai

#### 3.2.5.1 Le critère de délai par type d'application

Suivant le type de l'application, la définition du critère de délai et les exigences sur ce paramètre sont différentes. Pour les applications temps réel, le critère du délai représente

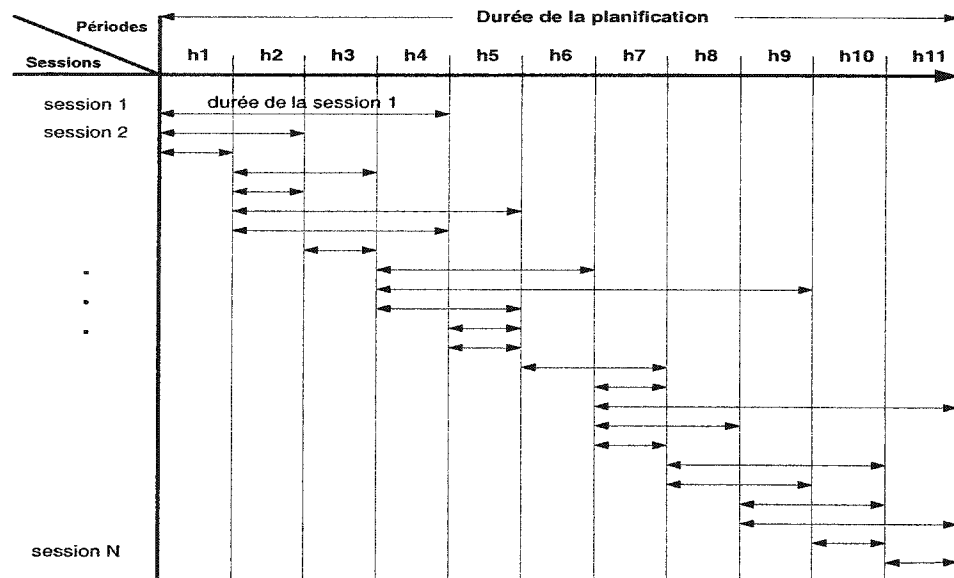


FIG. 3.6 – *Séquencement temporel: les débuts et fins de sessions correspondent avec les périodes*

le temps de transit d'un paquet de l'origine à sa destination et le temps d'inter-arrivée des paquets à l'arrivée. Pour le flux vidéo plus particulièrement, la priorité est mise sur le temps d'inter-arrivée. Ces deux paramètres fixent la qualité de la communication. Aussi, ces applications ont des exigences strictes pour ces paramètres. Il n'y a donc pas de retransmissions en cas d'erreur pour privilégier des délais courts.

Des applications non temps réel, comme la navigation Internet, ont des exigences un peu plus souples sur le délai, qui, dans ce cas, concerne le temps entre la requête et le téléchargement de la page. Le courrier électronique a des exigences encore moins strictes sur le délai. En effet, le destinataire n'attend pas de réponse avant un certain temps. Le délai, dans ce cas, est le temps entre l'émission et la réception totale du courrier.

Pour la navigation Internet ou le courrier électronique, les exigences sur le délai sont plus souples car l'intégralité des données est prioritaire. Aussi les délais sont pénalisés par des éventuelles retransmissions qui peuvent avoir lieu sur les liens radio comme sur les liens câblés. Les protocoles de retransmission ont ainsi une influence sur la qualité de service. On néglige les délais de retransmission sur les liens câblés car ces derniers sont beaucoup plus fiables que les liens radio. En effet, les interférences de puissance, la pseudo orthogonalité des codes, les pertes de propagation et les évanouissements rendent la liaison radio peu

sûre. De plus, le délai de transmission sur un lien radio est bien plus important que sur un lien câblé, les retransmissions sont donc plus pénalisantes sur les liens radio. Le tableau 3.4 résume les exigences sur le délai par application.

TAB. 3.4 – *Exigences sur le délai par application*

Applications	Définition du délai	Exigences sur le délai	Retransmission
Voix Vidéo conférence	Temps de transit d'un paquet et temps d'inter-arrivée des paquets à l'arrivée	Stricte Préservation des relations temporelles entre les entités	non
Flux video	Temps d'inter-arrivées des paquets	Stricte Préservation des relations temporelles entre les entités	non
Navigation Internet	Temps entre la requête et le chargement de la page	Souple Préservation de la charge	oui
Courrier électronique	Temps entre l'émission et la réception du courrier	Très souple Préservation de la charge	oui

### 3.2.5.2 Modélisation des containtes sur le délai

Seuls les temps de transit des paquets de l'origine à la destination sont modélisés. Le critère du délai apparaît sous forme de bornes maximales sur ces temps de transit. Une application souple sur le délai aura une borne plus large qu'une application exigeante.

Pour respecter ces bornes maximales, on impose des débit minimaux aux différents flots d'une session, aini les ressources disponibles testées par le CAC devront être suffisantes pour satisfaire au moins ces débits minimaux. Ces débits minimaux ne pourront être appliqués qu'aux applications VBR, car la procédure aura à choisir un débit pour les flots de la session. Dans le cas des applications CBR, la valeur de débit des flots est fixée par le type de codeur utilisé.

Le délai de transit d'un paquet d'une session  $s$  dans le sens montant (respectivement descendant), noté  $D_s^{UL}$  (respectivement  $D_s^{DL}$ ), est la somme des délais accumulés sur la partie câblée (WL) et sur la partie radio (RL) du réseau.

$$D_s^{UL} = D_{f^{UL-WL}} + D_{f^{UL-RL}}, \quad (3.19)$$

$$D_s^{DL} = D_{f^{DL-WL}} + D_{f^{DL-RL}}. \quad (3.20)$$

### Délai sur les liens câblés

Le délai subi par un paquet sur les liens câblés est dû au délai de transmission et au délai de propagation sur les liens. A cela s'ajoute un délai d'attente dans les files WFQ et un délai de traitement au niveau des routeurs :

$$D_{f_{UL-WL}} = D_{f_{UL-WL}}^{trans} + D_{f_{UL-WL}}^{propag} + D_{f_{UL-WL}}^{trait} + D_{f_{UL-WL}}^{attente}, \quad (3.21)$$

$$D_{f_{DL-WL}} = D_{f_{DL-WL}}^{trans} + D_{f_{DL-WL}}^{propag} + D_{f_{DL-WL}}^{trait} + D_{f_{DL-WL}}^{attente}. \quad (3.22)$$

– Délai de transmission.

Le transport par paquets IP intervient lors de l'évaluation du délai de transmission quand on prend comme référence pour la taille d'un paquet la taille caractéristique d'un paquet dans un réseau IP pour l'application donnée.

D'après la section 3.2.2, le débit minimal sur les liens câblés est la valeur du débit  $r_t^{UL}$  du RAB  $r^{UL}$  sélectionné pour le sens montant et  $\mathcal{R}(r_t^{DL})$  pour le sens descendant. Si  $m_{f_{UL}}$  et  $m_{f_{DL}}$  sont les nombres de liens câblés sur les routes des flots montant et descendant, alors les délais maximaux de transmission se calculent par :

$$D_{f_{UL-WL}}^{max-trans} = m_{f_{UL}} \times \frac{Taille\_paquet\_UL}{r_t^{UL}}, \quad (3.23)$$

$$D_{f_{DL-WL}}^{max-trans} = m_{f_{DL}} \times \frac{Taille\_paquet\_DL}{\mathcal{R}(r_t^{DL})}. \quad (3.24)$$

– Délai de propagation.

Le délai de propagation est le délai de propagation du bit sur le lien, il dépend de la longueur du lien. Pour simplifier on prendra une valeur moyenne de ce délai.

$$D_{f_{UL-WL}}^{propag} = m_{f_{UL}} \times Temps\_propagation/liens, \quad (3.25)$$

$$D_{f_{DL-WL}}^{propag} = m_{f_{DL}} \times Temps\_propagation/liens. \quad (3.26)$$

– Délai de traitement.

Le paquet doit être traité par le routeur pour être correctement routé et pour être classé dans la file d'attente de sa connexion. Le contrôle d'erreur effectué sur chaque

routeur implique aussi un délai que l'on prend en compte dans le délai de traitement. Le nœud d'origine n'est pas considéré comme un routeur alors que le nœud de destination, qui est la station de base dans le sens descendant, sera comptabilisé comme un routeur avec un délai de traitement. Le nombre de routeurs est donc égal au nombre de liens câblés sur la route :

$$D_{f_{UL},wl}^{trait} = m_{f_{UL}} \times Temps\_traitement/routeur, \quad (3.27)$$

$$D_{f_{DL},wl}^{trait} = m_{f_{DL}} \times Temps\_traitement/routeur. \quad (3.28)$$

— Délai d'attente.

Au niveau de chaque routeur, le paquet est classé dans la file d'attente attitrée à la session. Avec la variante RPPS de la discipline WFQ, décrite à la section 3.1.2, le délai d'attente d'un paquet accumulé dans les files des routeurs depuis l'origine jusqu'à la destination est majoré par la borne du pire cas  $D_s^*$  de la formule (3.4).

Dans le sens montant, on impose que le débit du flot sur les liens câblés soit supérieur ou égal au débit du RAB alloué sur le lien radio pour éviter toute accumulation à la station de base (équation (3.16)). Ainsi, on réserve le maximum de bande passante sur les liens câblés car le débit du RAB est le débit maximal auquel les données arrivent à la station de base. En effet, le lien radio rend la source conformante car son débit crête est majoré par le débit du RAB, il joue en quelque sorte le rôle d'un filtre en seau percé dont le taux de remplissage du seau serait la valeur de débit du RAB et la profondeur du seau serait nulle. Pour cette raison, il n'y a pas besoin de filtre dans le sens montant au niveau de la station de base (figure 3.1).

Quand la profondeur de seau est nulle ( $\sigma = 0$ ), la source n'est pas autorisée à émettre des paquets en rafales à un débit supérieur au taux de remplissage. Avec WFQ, le taux minimal garanti de service au niveau des routeurs est supérieur au taux de remplissage du seau (propriété 3.3). Ce dernier est fixé à la valeur du débit du RAB qui, d'après les remarques précédentes, représente déjà la valeur maximale du taux d'arrivée des paquets sur les liens câblés. Il n'y a donc pas de délai d'attente dans les files des routeurs pour le sens montant.

$$D_{f_{UL-WL}}^{attente} = 0. \quad (3.29)$$

En revanche, dans le sens descendant, le paquet subit un délai d'attente dans les routeurs. En effet, dans ce sens la source est rendue conformante par un filtre en sceau percé placé au niveau du nœud d'accès externe dont le taux de remplissage n'est plus la valeur maximale de débit possible pour le flot mais plutôt sa valeur déclarée. Les sources pourront envoyer des paquets en rafales et de ce fait, les délais d'attente dans les files ne sont plus nuls. Mais comme les sources sont rendues conformantes, on peut utiliser la borne (3.4).

$$D_{f_{DL-WL}}^{attente} \leq D_s^*. \quad (3.30)$$

#### Délai sur les liens radio

Le paquet IP dans le sens descendant est scindé par la station de base pour être transmis par trames sur le lien radio. Dans le sens montant, plusieurs trames reçues sont réorganisées pour former un paquet IP. Le délai de transmission du paquet sur un lien radio correspond donc au délai pour transmettre, ou recevoir suivant le sens, toutes les trames impliquées dans un paquet IP.

La durée d'une trame est fixe (20 ms pour CDMA2000), le nombre de bits transmis sur une trame dépend donc du débit du RAB sélectionné. Par conséquent, le nombre de trames nécessaires pour transmettre le paquet IP dépend aussi du débit du RAB sélectionné.

Si  $N_{r_t^{UL}}$  et  $N_{r_t^{DL}}$  sont les nombres de trames nécessaires pour transmettre un paquet IP aux débits  $r_t^{UL}$  et  $r_t^{DL}$  sur les liens radio montant et descendant, il vient:

$$N_{r_t^{UL}} = \lfloor \frac{8 \times Taille\_paquet\_UL(octets)}{Duree\_trame \times r_t^{UL}} \rfloor, \quad (3.31)$$

$$N_{r_t^{DL}} = \lfloor \frac{8 \times Taille\_paquet\_DL(octets)}{Duree\_trame \times r_t^{DL}} \rfloor. \quad (3.32)$$

Le délai de transmission d'une trame est d'en moyenne 40 ms pour une application temps réel et 100 ms pour une application non temps réel. Ce délai peut être pénalisé d'un délai supplémentaire si des retransmissions sont nécessaires. Le nombre moyen de retransmissions dépend du taux d'erreurs associé au RAB sélectionné. Le protocole de retransmission



modélisé est la répétition sélective ou SRP. L'annexe I décrit le principe de ce protocole ainsi que le détail des calculs pour le délai de transmission moyen d'un trame en tenant compte des retransmissions. Le tableau suivant récapitule les résultats de l'annexe I.

TAB. 3.5 – *Délai moyen de transmission d'une trame sur un lien radio*

FER	$D_{\text{FER}}^{\text{transmission-radio}}$
0.5 %	101.01 ms
1 %	102.03 ms
2 %	104.11 ms
5 %	110.68 ms
10 %	122.78 ms

Le délai de transmission d'un paquet IP sur un lien radio est alors:

$$D_{f_{\text{UL,RL}}}^r = N_{r_t^{\text{UL}}} \times D_{\text{FER}}^{\text{transmission-radio}}, \quad (3.33)$$

$$D_{f_{\text{DL,RL}}}^r = N_{r_t^{\text{DL}}} \times D_{\text{FER}}^{\text{transmission-radio}}. \quad (3.34)$$

### 3.2.5.3 Informations caractéristiques d'une session

Les informations qui caractérisent une session sont :

- l'application de la session,
- la priorité de l'utilisateur,
- la position de l'utilisateur,
- le début de la session,
- la durée de la session,
- l'ensemble des flots qui lui sont associés,
- la route de la session.

Les applications VBR ont en plus les caractéristiques suivantes :

- les débits maximal et minimal pour chaque flot,
- la taille moyenne d'un paquet (délai),
- la taille maximale d'un paquet (délai),
- la profondeur du filtre en seau percé placé au niveau du nœud d'accès externe (délai),
- les bornes sur le délai.

### 3.3 Limites du modèle

#### 3.3.1 Modèle anticipatif

Le modèle est dit anticipatif car le contrôle d'admission accepte ou refuse une session sur la base des tests de disponibilité des ressources mais aussi sur la base des sessions antérieures et surtout postérieures. Ceci est possible car le modèle n'est pas itératif mais global. Il a toutes les informations sur les requêtes à venir. Ce type de contrôle d'admission n'a pas de réalité physique. Les contrôles d'admission implémentables testent les ressources disponibles au moment de la requête en ne considérant que les sessions en cours. Si les tests sont positifs alors la session est forcément acceptée. Il n'y a pas de pronostics faits sur d'éventuelles sessions à venir plus prioritaires. On nomme ces contrôles d'admission causal. L'avantage du modèle anticipatif est qu'il fournit une meilleure solution que si le contrôle d'admission était causal. En effet, le modèle anticipatif va pouvoir répartir dans le temps les sessions refusées pour optimiser les ressources tout en respectant les taux de services, ce que n'est pas autorisé à faire le contrôle d'admission causal.

La solution fournie par le modèle anticipatif peut donc être considérée comme une borne pour tester l'efficacité de différentes stratégies de contrôle d'admission implémentées dans la réalité.

#### 3.3.2 Imprécisions sur le dimensionnement dues au séquençement temporel

Le séquençement temporel, décrit à la section 3.2.4, entraîne une surestimation des durées des sessions. En effet, la durée d'une session est transformée en une valeur discrète juste supérieure pour coïncider avec les périodes. Plus les périodes sont courtes, plus cette surestimation est réduite. Cependant, les nombres de variables et de contraintes seront aussi plus importants. Il faudra trouver un compromis.

De plus, on suppose que la durée d'une session est connue et constante quelle que soit le débit sélectionné par la procédure d'optimisation. Cette hypothèse n'est pas restrictive pour les applications temps réel comme la voix, la vidéo conférence et le flux vidéo car la durée de la connexion dépend uniquement du comportement de l'utilisateur. Pour les applications non temps réel comme la navigation Internet et le courrier électronique, la durée de la connexion dépend du temps de téléchargement des données. La durée réelle dépend donc

du débit de service et ne correspond plus forcément à la durée déclarée. Cela entraîne une surestimation ou une sousestimation du dimensionnement. En effet, les paramètres de routage d'une session indiquent quels liens câblés du réseau de cœur elle utiliserait si elle était acceptée et sur quelles périodes. Les ressources sont donc réservées sur la base de la durée déclarée entraînant ainsi un sur-dimensionnement si la session se finit réellement plus tôt et un sous-dimensionnement si la session se termine réellement plus tard.

Une façon de résoudre ce problème serait de considérer plusieurs scénarios pour lesquels la durée d'une session dépendrait du débit choisi. Les paramètres de routage deviendraient alors des variables rendant le modèle bien plus complexe.

### 3.3.3 Les formules de capacités radio

Récapitulons les hypothèses limitant la précisions des formules de capacités radio :

- le contrôle de puissance est supposé parfait. Les puissances des sessions sont donc toujours ajustées,
- les puissances liées aux signaux pilotes et aux signaux de synchronisation ne sont pas prises en compte dans le sens montant,
- le trafic par type d'application est supposé uniforme sur toutes les cellules. Ainsi, on peut écrire que la puissance fournie par toutes les stations de base est la même,
- les cellules ne sont pas sectorisées,
- le modèle d'atténuation est simplifié. Il ne tient pas compte des évanouissements de Rayleigh et des évanouissements liés à des réflexions partielles sur les obstacles.

### 3.3.4 Le délai pour les sessions en *soft handover*

Le modèle prend en compte, dans les formules de capacité radio, la réduction du SIR requis pour une session en *soft handover*. Cette réduction se manifeste sous la forme d'un gain ( $SHG^{UL}$  et  $SHG^{DL}$ ) sur le SIR associé au RAB cible. D'après la section 3.1.4, on sait que cette réduction entraîne un taux d'erreur plus important. La qualité du lien est alors moins bonne. Le FER du RAB effectif sera donc différent du FER du RAB cible, c'est-à-dire que le flot sera bien servi au débit du RAB cible mais le FER réel sera plus important. Cet effet n'est pas pris en compte dans le modèle où on raisonne comme si le RAB était

inchangé. Cette hypothèse a un impact uniquement sur l'évaluation du délai. En effet, le délai sur le lien radio dépend du FER (formules (3.33) et (3.34)). En considérant le FER du RAB cible dans ces formules, le délai pour les applications VBR sera alors sous-évalué. D'après le tableau 3.5, on peut voir que, pour une trame, la sous-évaluation se mesure en unités de millisecondes si le FER réel double. Ce qui devient des dizaines de millisecondes pour un paquet IP en considérant que 10 trames suffisent pour découper un paquet IP imposant, ce qui n'est donc pas négligeable.

### 3.4 Stratégie de dimensionnement

L'objectif du dimensionnement est de :

- minimiser les capacités des liens câblés pour la partie câblée (WL),
- minimiser le nombre de stations de base et identifier leurs positions pour la partie radio (RL).

Les liens câblés ont, sur chaque période, une certaine capacité occupée par les sessions qui les traversent. Le dimensionnement de la partie câblée consiste classiquement à prendre comme capacité optimale d'un lien la capacité maximale de ces capacités occupées sur toutes les périodes.

Pour le dimensionnement de la partie radio, la stratégie est plus originale. Au départ, nous considérons une aire géographique à couvrir, associée à une topologie d'un réseau câblé. Le nombre de BSC est fixé par la topologie et on divise la surface en plusieurs surfaces équivalentes qui seront couvertes, chacune, par un BSC. On répartit ensuite de façon uniforme un certain nombre de stations de base potentielles. Chaque station de base sera reliée au BSC qui couvre la surface où elle se trouve. On crée volontairement un sur-dimensionnement initial avec un nombre élevé de stations de base potentielles. On s'assure ainsi que la procédure d'optimisation trouvera une solution si le trafic est élevé. Les stations de base sont dites potentielles car la procédure d'optimisation choisit lesquelles sont nécessaires, dans une configuration optimale, pour supporter la charge de trafic en respectant les exigences de qualité de service des sessions et les niveaux de service globaux (GoS). La procédure est globale et va considérer toutes les périodes simultanément pour accepter ou refuser les sessions. Une station de base potentielle devient une station de base nécessaire au

dimensionnement si elle a été sélectionnée au moins une fois pour une période quelconque. La procédure utilise les formules de capacités radio pour déterminer quelle station de base prendra en charge l'utilisateur mobile. Les positions des deux entités interviennent dans le choix de la procédure à travers les coefficients d'atténuation  $a_{s,0}$  et  $a_{s,i}$  dans la formule de capacité radio 3.13 (sens descendant). Ainsi, à l'image des modèles de dimensionnement radio<sup>[17], [18]</sup> et <sup>[19]</sup>, décrits à la section 2.1, notre stratégie de dimensionnement se base sur les coefficients d'atténuation pour sélectionner les positions des stations de base nécessaires.

Le générateur de trafic répartit uniformément sur la surface à dimensionner les sessions d'une même application et commençant à la même période. Les positions des utilisateurs mobiles sont donc connues. Par rapport à cette position, on définit un ensemble de stations de base potentielles pour l'utilisateur mobile, forçant la procédure à choisir une ou deux stations de base dans cet ensemble. De plus, on définit dans cet ensemble un sous-ensemble qui contient les stations de base potentielles qui pourront supporter la session, seule ou en *soft* handover. Les stations de base potentielles, qui ne font pas partie de ce sous-ensemble, pourront supporter la session mais sans *soft* handover. Les stations de base potentielles sont ainsi classées suivant un critère de distance. En effet, une station de base potentielle trop éloignée de l'utilisateur mobile ne pourra pas physiquement le prendre en charge. De même, une station de base potentielle trop près de l'utilisateur mobile ne pourra pas le prendre en *soft* handover. En classant ainsi les stations de base potentielles, on réduit le choix de la procédure et on limite le nombre de variables et le nombre de contraintes.

Les figures 3.7 et 3.8 illustrent la stratégie de dimensionnement radio.

### 3.5 Limites de la stratégie

#### 3.5.1 Distribution uniforme

Les terminaux mobiles qui font une requête pour la même application et commençant à la même période sont répartis uniformément sur la surface géographique à dimensionner. Avec la répartition uniforme des stations de base potentielles, on s'approche alors de l'hypothèse d'un trafic d'intensité homogène par période sur toutes les cellules. Cette hypothèse est nécessaire aux formules de capacité radio (sections 3.1.5.3 et 3.1.5.4).

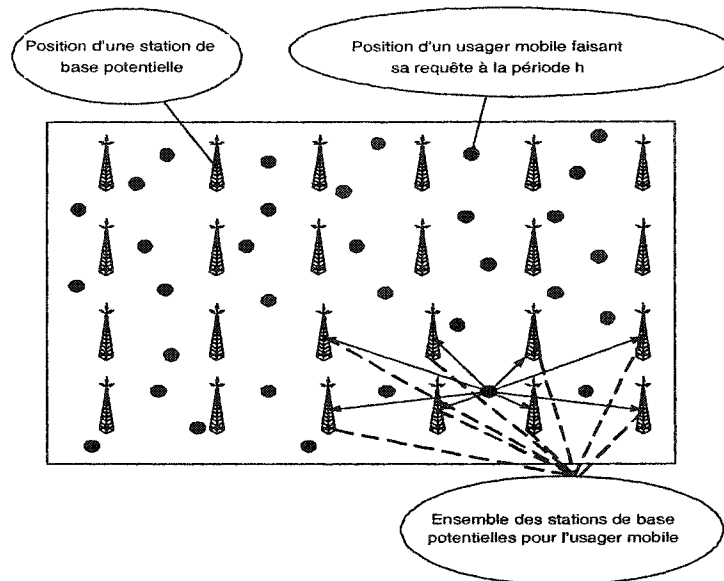


FIG. 3.7 – Répartition uniforme par type d'application des sessions commençant à la période  $h$

Cependant, comme la procédure peut refuser des sessions et qu'elle ne sélectionne pas toutes les stations de base potentielles, il n'est pas garanti que l'intensité reste homogène pour chaque période. La puissance fournie et reçue par une station de base n'est alors plus la même pour toutes les stations de base. Cependant, on peut prédire que la procédure d'optimisation va répartir les stations de base sélectionnées à peu près uniformément pour bien couvrir toute la région. De plus, les taux de blocage sont suffisamment faibles pour que le trafic accepté dans le réseau reste proche du trafic initial et donc que l'intensité du trafic restera homogène en moyenne.

### 3.5.2 Les BSCs et les stations de base potentielles pour un usager mobile

D'après la section 3.2.3, le BSC lié à une session est prédéterminé ainsi que ses routes montante et descendante. C'est du mono-routage statique : chaque flot a une seule route déterminée par avance et celle-ci ne change pas durant toute la connexion. Ceci implique que la procédure d'optimisation doit sélectionner une station de base (ou deux en cas de *soft handover*) reliée au même BSC que la session. Sinon le routage deviendrait du multi-routage. Les routes de la session ne pourraient plus être déterminées par avance, les liens câblés du réseau de cœur étant devenus potentiels suivant le BSC sélectionné. En plus de

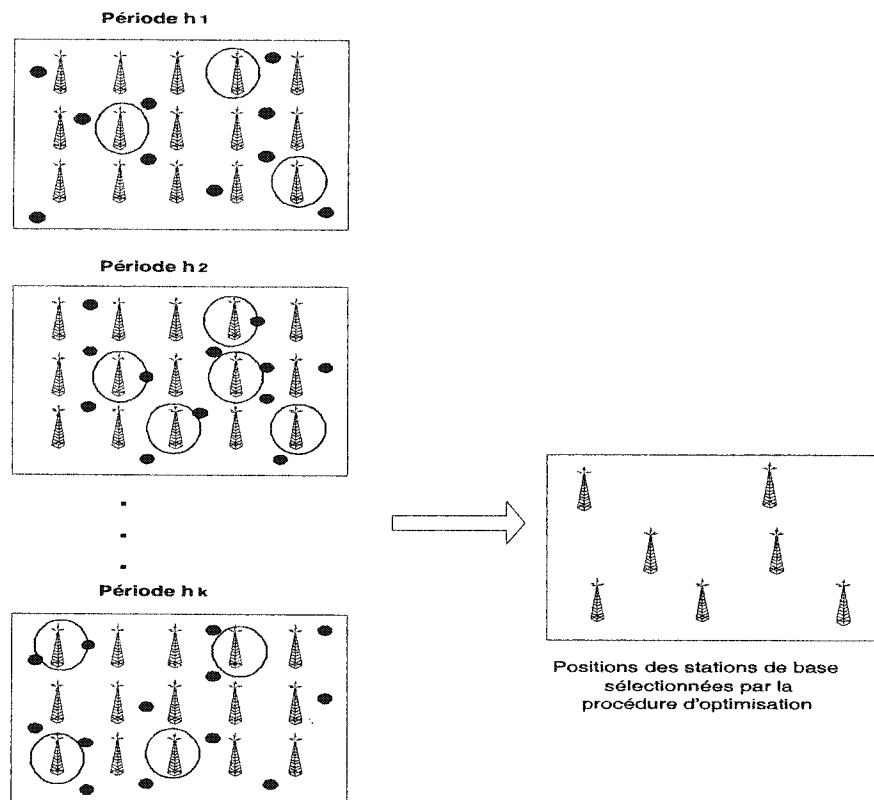


FIG. 3.8 – Sélection des stations de base à installer physiquement pour supporter le trafic sur toutes les périodes

choisir les stations de base, la procédure aurait alors le choix entre plusieurs routes reliées à des BSC différents.

Le choix du mono-routage a été fait pour limiter la complexité du modèle. Mais le multi-routage est opéré dans les réseaux réels. La modélisation du multi-routage sera donc une opportunité d'améliorer le modèle.

### 3.6 Récapitulatif des points principaux de la modélisation

- l'architecture du réseau est divisée en deux parties:
  - le réseau de cœur,
  - le réseau d'accès radio (RAN).
- le modèle proposé fournit un outil de dimensionnement global pour le réseau de cœur et le réseau d'accès radio,

- un filtre en seau percé est placé pour chaque session dans son nœud d'accès externe pour contrôler la source dans le sens descendant,
- le lien radio joue le rôle d'un filtre en seau percé de profondeur nulle. Les sources sont donc rendues conformantes dans le sens montant dès leur entrée dans le réseau câblé,
- la politique de service des paquets sur les liens câblés est WFQ,
- la génération du trafic est déterministe,
- le temps est découpé en périodes élémentaires,
- les connexions sont toujours entre un usager mobile (terminal mobile) et un usager fixe (réseaux externes),
- les routes sont prédéterminées et elles restent les mêmes durant toute la connexion. C'est du mono-routage statique,
- une session interactive est modélisée par quatre flots :  $f^{UL-WL}$ ,  $f^{DL-WL}$ ,  $f^{UL-RL}$  et  $f^{DL-RL}$ ,
- une session non interactive est modélisée par deux flots :  $f^{UL-WL}$  et  $f^{UL-RL}$  ou  $f^{DL-WL}$  et  $f^{DL-RL}$ ,
- pour les applications CBR, les débits des flots sur les liens câblés et sur les liens radio sont fixés,
- pour les applications VBR, les débits sont choisis par la procédure d'optimisation,
- pour ces applications, on impose que le débit du flot  $f^{UL-WL}$  soit supérieur ou égal au débit du flot  $f^{UL-RL}$  (3.16),
- pour ces applications, on impose que le débit du flot  $f^{DL-RL}$  soit supérieur ou égal au débit du flot  $f^{DL-WL}$  (3.17),
- pour ces applications, des débits minimaux sont imposés pour garantir les bornes sur le délai,
- il n'y a pas de délai d'attente dans les routeurs pour les flots montant,
- le délai d'attente pour les flots descendant est majoré par la borne de Parekh et Gallager (3.4),
- grâce au contrôle d'admission, une session acceptée est assurée d'avoir les ressources nécessaires sur les liens câblés et sur les liens radio pour satisfaire ses exigences de qualité de service,
- le contrôle d'admission est anticipatif et tient compte des priorités des usagers,



- un terminal mobile peut être pris en *soft* handover par deux stations de base,
- les deux stations de base impliquées dans le *soft* handover sont reliées au même BSC,
- les deux stations de base fournissent le même service pour chaque flot,
- la procédure peut refuser des sessions tant que les taux minimaux d'acceptation (GoS) sont respectés,
- les stations de base sont initialement potentielles,
- la répartition initiale des stations de base potentielles est uniforme et l'intensité du trafic est homogène sur toutes les cellules à chaque période,
- si une station de base est activée au moins une fois pour une période, alors elle sera sélectionnée par la procédure,
- le dimensionnement résultant fournit, pour un trafic donné, les capacités optimales des liens câblés et le nombre optimal des stations de base nécessaires ainsi que leurs positions,
- la mobilité des usagers n'est pas prise en compte dans les formules de capacités radio ni au niveau du routage.

## CHAPITRE 4

### FORMULATION MATHÉMATIQUE DU MODÈLE D'OPTIMISATION

Un modèle d'optimisation se compose, d'une part, d'une fonction objectif qui définit le coût en fonction de certaines variables et d'autre part d'un ensemble de contraintes impliquant l'ensemble des variables. L'optimisation consiste à trouver une valeur pour chacune de ces variables telle que le coût soit minimal ou maximal suivant le problème et que toutes les contraintes soient respectées simultanément. Dans notre étude, l'objectif est de réduire le coût du dimensionnement, c'est donc un problème de minimisation. La stratégie du dimensionnement, expliquée à la section 3.4, conditionne la forme de la fonction objectif alors que les modélisations du réseau et du trafic sont appliquées à travers les contraintes.

#### 4.1 Notations

##### 4.1.1 Notations générales

$S$	ensemble des sessions,
$s$	une session,
$P$	ensemble des priorités pour les usagers (Gold et Silver),
$A$	ensemble des différentes applications,
$S^p$	ensemble des sessions de priorité $p$ ,
$S^a$	ensemble des sessions d'application $a$ ,
$S^{\text{CBR}}$	ensemble des sessions dont l'application est CBR ,
$S^{\text{VBR}}$	ensemble des sessions dont l'application est VBR ,
$F$	ensemble des flots,
$f$	un flot,
$F^{\text{UL-WL}}$	ensemble des flots sur les liens câblés pour le sens montant,
$F^{\text{DL-WL}}$	ensemble des flots sur les liens câblés pour le sens descendant,
$F^{\text{RL}}$	ensemble des flots sur les liens radio,

$F^{\text{UL-RL}}$	ensemble des flots sur les liens radio pour le sens montant,
$F^{\text{DL-RL}}$	ensemble des flots sur les liens radio pour le sens descendant,
$H$	ensemble des périodes de temps,
$h$	une période.

#### 4.1.2 Notations pour les niveaux de service et la qualité de service

$\text{Br}_{a,p}$	taux de blocage pour les sessions d'application $a$ et de priorité $p$ ,
$Q_{r,a,p}$	coefficient de qualité de service défini pour les sessions d'application $a$ , de priorité $p$ et servies au RAB $r$ .

#### 4.1.3 Notations pour les sessions

Pour une session  $s$  donnée :

$f_s^{\text{UL-WL}}$	flot dans le sens montant sur les liens câblés de la route montante de la session $s$ ,
$f_s^{\text{DL-WL}}$	flot dans le sens descendant sur les liens câblés de la route descendante de la session $s$ ,
$f_s^{\text{UL-RL}}$	flot dans le sens montant sur les liens radio associés à la session $s$ ,
$f_s^{\text{DL-RL}}$	flot dans le sens descendant sur les liens radio associés à la session $s$ ,
$F_s$	ensemble des quatre flots, définis ci-dessus, associés à la session $s$ ,
$F_s^{\text{WL}}$	ensemble des deux flots définis sur les liens câblés associés à la session $s$ ,
$F_s^{\text{RL}}$	ensemble des deux flots définis sur les liens radio associés à la session $s$ ,
$T_s^{\text{UL}}$	ensemble des RABs proposés à la session $s$ dans le sens montant,
$T_s^{\text{DL}}$	ensemble des RABs proposés à la session $s$ dans le sens descendant,
$\sigma_s$	gigue tolérée dans le sens descendant pour la session $s$ ,
$L_s^{\text{BS}}$	ensemble des stations de base potentielles qui peuvent prendre en charge la session $s$ ,

$L_s^{\text{BS,SH}}$  sous-ensemble de  $L_s^{\text{BS}}$  qui contient les stations de base potentielles qui peuvent éventuellement prendre la session  $s$  en *soft* handover.

Pour les flots associés à la session  $s$  :

$L_f^{\text{max}}$	taille maximale d'un paquet du flot $f$ de l'ensemble $F_s^{\text{WL}}$ ,
$L_f$	taille moyenne d'un paquet du flot $f$ de l'ensemble $F_s^{\text{WL}}$ ,
$m_f$	nombre de liens câblés sur la route du flot $f$ de l'ensemble $F_s^{\text{WL}}$ ,
$\bar{\rho}_f$	débit maximal du flot $f$ de l'ensemble $F_s^{\text{WL}}$ quand $s$ est de type VBR,
$\underline{\rho}_f$	débit minimal du flot $f$ de l'ensemble $F_s^{\text{WL}}$ quand $s$ est de type VBR,
$\rho_f$	débit déclaré du flot $f$ de l'ensemble $F_s^{\text{WL}}$ quand $s$ est de type CBR,
$\bar{t}_f$	débit maximal du flot $f$ de l'ensemble $F_s^{\text{RL}}$ quand $s$ est de type VBR,
$\underline{t}_f$	débit minimal du flot $f$ de l'ensemble $F_s^{\text{RL}}$ quand $s$ est de type VBR,
$t_f$	débit déclaré du flot $f$ de l'ensemble $F_s^{\text{RL}}$ quand $s$ est de type CBR,
$D_s^{\text{DL-max}}$	délai maximal de transit de bout en bout d'un paquet du flot $f_s^{\text{DL-WL}}$ ,
$D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t}$	délai maximal de transmission sur les liens câblés d'un paquet du flot $f_s^{\text{DL-WL}}$ quand le RAB $r$ est sélectionné,
$D_{f_s^{\text{DL-WL}}}^{\text{propag}}$	délai de propagation pour un paquet du flot $f_s^{\text{DL-WL}}$ ,
$D_{f_s^{\text{DL-WL}}}^{\text{trait}}$	délai de traitement dans les routeurs pour un paquet du flot $f_s^{\text{DL-WL}}$ ,
$D_{f_s^{\text{DL-RL}, r_t}}$	délai de transmission sur les liens radio pour un paquet du flot $f_s^{\text{DL-RL}}$ quand le RAB $r$ est sélectionné sur le lien radio.

Pour chaque session  $s$ , des paramètres de routage sont associés aux flots de l'ensemble  $F_s^{\text{WL}}$  :

$$a_{\ell h f} = \begin{cases} 1 & \text{si le flot } f \text{ de la session traverse le lien câblé } \ell \text{ pendant la période } h, \\ 0 & \text{sinon.} \end{cases}$$

$$b_{ihf} = \begin{cases} 1 & \text{si le flot } f \text{ de la session peut potentiellement aller jusqu'à la station de base} \\ & i \text{ pendant la période } h. b_{ihf} \text{ est défini pour les stations de base potentielles} \\ & i \text{ tel que } i \in L_s^{\text{BS}}, \\ 0 & \text{sinon.} \end{cases}$$

#### 4.1.4 Notations pour les liens câblés

$m$  nombre total de liens câblés dans le réseau de cœur,  
 $d_\ell$  longueur du lien câblé  $\ell$ .

#### 4.1.5 Notations pour les liens radio

$n_{\text{BS}}$  nombre de stations de base potentielles égal au nombre de liens câblés  
du réseau d'accès radio,  
 $T$  ensemble des RABs disponibles sur un lien radio,  
 $T^{\text{UL}}$  ensemble des RABs disponibles sur un lien radio dans le sens montant,  
 $T^{\text{DL}}$  ensemble des RABs disponibles sur un lien radio dans le sens descendant,  
 $r_t$  débit du RAB  $r$ ,  
 $r_{\text{FER}}$  FER du RAB  $r$ ,  
 $r^{\text{UL}}$  un RAB dans le sens montant,  
 $r^{\text{DL}}$  un RAB dans le sens descendant.

### 4.2 Définition des variables

Notre problème d'optimisation est mixte. Il contient des variables continues, comprises entre une borne maximale et une borne minimale, et des variables binaires, qui ne peuvent prendre que les valeurs 0 ou 1.

Pour chaque session, une variable continue représentant le débit est associée à chaque flot de la session :

- $x_f$  : variable continue pour le débit du flot  $f$ .

$x_{f_s^{UL-WL}}$  = débit du flot montant uniquement sur les liens câblés du réseau de coeur,

$x_{f_s^{UL-RL}}$  = débit du flot montant uniquement sur les liens radio,

$x_{f_s^{DL-WL}}$  = débit du flot descendant uniquement sur les liens câblés du réseau de coeur,

$x_{f_s^{DL-RL}}$  = débit du flot descendant uniquement sur les liens radio.

Comme le lien entre le BSC et la station de base potentielle dépend du choix de la procédure d'optimisation (section 3.2.3), on définit des variables spéciales pour les débits des flots sur ces liens.

- $x_f^i$  : variable continue pour le débit du flot  $f$  ( $\in F_s^{WL}$ ) sur le lien câblé du réseau d'accès radio, reliant la station de base  $i$  à son BSC.

$$x_{f^{WL}}^i = \begin{cases} x_{f^{WL}} & \text{si la station de base } i \text{ qui appartient à l'ensemble } L_s^{BS} \\ & \text{est sélectionnée pour la session } s, \\ 0 & \text{sinon.} \end{cases}$$

- $y_s$  : variable binaire pour le CAC. Une variable de CAC est définie pour chaque session  $s$ .

$$y_s = \begin{cases} 1 & \text{uniquement si tous les flots } f \in F_s \text{ sont acceptés,} \\ 0 & \text{sinon.} \end{cases}$$

Grâce à cette variable, la requête d'une session sera acceptée ou refusée sur la base des ressources disponibles pour chacun de ses flots.

- $\alpha_{irf}$  : variable binaire pour sélectionner un RAB  $r$  sur le lien radio, reliant le terminal mobile à la station de base  $i$ .  $\alpha_{irf}$  est définie pour les stations de base  $i$  tel que  $i \in L_s^{BS}$ , pour les RABs  $r$  tel que  $r \in T_s^{UL}$  si  $f = f_s^{UL-RL}$  et  $r \in T_s^{DL}$  si  $f = f_s^{DL-RL}$ .

$$\alpha_{irf} = \begin{cases} 1 & \text{si la station de base } i \text{ et le RAB } r \text{ sont sélectionnés pour le flot } f, \\ 0 & \text{sinon.} \end{cases}$$

- $C_\ell$  : variable continue pour la capacité du lien câblé  $\ell$  du réseau de cœur.
- $C_{\ell_{BS_i}}$  : variable continue pour la capacité du lien câblé reliant la station de base  $i$  à son BSC,  $C_{\ell_{BS_i}} = 0$  si la station de base  $i$  n'est pas sélectionnée.
- $x^{BS_i}$  : variable binaire pour sélectionner la station de base  $i$ .

$$x^{BS_i} = \begin{cases} 1 & \text{si la station de base } i \text{ a été choisie au moins une fois pour} \\ & \text{prendre en charge une session,} \\ 0 & \text{sinon.} \end{cases}$$

Les variables suivantes modélisent le *soft* handover. Elles sont définies pour chaque session  $s$  :

- $z_{fr}$  : variable binaire pour le *soft* handover.  $z_{fr}$  est définie pour les RABs  $r$  tel que  $r \in T_s^{\text{UL}}$  si  $f = f_s^{\text{UL-RL}}$  et  $r \in T_s^{\text{DL}}$  si  $f = f_s^{\text{DL-RL}}$ .

$$z_{fr} = \begin{cases} 1 & \text{si la session } s \text{ est en } \textit{soft} \text{ handover, avec } f \in F^{\text{RL}} \text{ et } r \text{ le RAB} \\ & \text{sélectionné,} \\ 0 & \text{sinon.} \end{cases}$$

- $w_{irf}$  : variable binaire pour le *soft* handover.  $w_{irf}$  est définie pour les stations de base  $i$  tel que  $i \in L_s^{\text{BS-SH}}$ , pour les RAB  $r$  tel que  $r \in T_s^{\text{UL}}$  si  $f = f_s^{\text{UL-RL}}$  et  $r \in T_s^{\text{DL}}$  si  $f = f_s^{\text{DL-RL}}$ .

$$w_{irf} = \begin{cases} 1 & \text{si la session } s \text{ est en } \textit{soft} \text{ handover avec la station de base } BS_i, \\ & \text{avec } f \in F^{\text{RL}} \text{ et } r \text{ le RAB sélectionné,} \\ 0 & \text{sinon.} \end{cases}$$

Ces dernières variables sont uniquement définies pour les stations de base qui appartiennent à l'ensemble  $L_s^{\text{BS-SH}}$ , c'est-à-dire pour les stations de base qui peuvent de façon réaliste prendre la session en *soft* handover.

### 4.3 Fonction objectif

L'objectif du dimensionnement est de minimiser les ressources matérielles, c'est à dire les liens câblés et les stations de base. En supposant que le coût associé à un lien câblé est proportionnel à sa longueur et à sa capacité, la fonction objectif s'écrit :

$$\min \left( w_c \sum_{\ell=1}^m d_{\ell} C_{\ell} + w_c \sum_{i=1}^{n_{BS}} d_{\ell_{BS_i}} C_{\ell_{BS_i}} + w_{BS} \sum_{i=1}^{n_{BS}} BS_i \right). \quad (4.1)$$

Où  $w_c$  est le coût par unité de longueur et par unité de capacité d'un lien câblé et  $w_{BS}$  le coût d'une station de base. Ces deux coefficients sont normalisés.

### 4.4 Contraintes

Toutes les contraintes sont linéaires. C'est un choix volontaire pour faciliter la résolution rendue déjà critique par le nombre élevé de variables binaires dans le modèle. Privilégier la linéarité des contraintes a coûté en nombre de variables et de contraintes, en effet, les variables  $w_{itf}$  auraient pû être évitées et certaines contraintes auraient pû être regroupées en contraintes quadratiques.

On a pu voir à la section 3.2.2.2 que l'attribution des flots dépend du type de l'application de la session. Certaines contraintes sont donc spécifiques des sessions de type VBR et d'autres des sessions de type CBR.

#### 4.4.1 Contrôle d'admission

Pour chaque session  $s$  avec son ensemble de flot  $F_s = \{f_s^{UL-WL}, f_s^{UL-RL}, f_s^{DL-WL}, f_s^{DL-RL}\}$ , les contraintes de CAC sont les suivantes.

**Si  $s$  est de type CBR :**

- sur les liens câblés :

$$x_{f_s^{UL-WL}} = y_s \rho_{f_s^{UL-WL}}. \quad (4.2)$$

$$x_{f_s^{DL-WL}} = y_s \rho_{f_s^{DL-WL}}. \quad (4.3)$$



– sur les liens radio :

$$x_{f_s^{UL-RL}} = y_s t_{f_s^{UL-RL}}. \quad (4.4)$$

$$x_{f_s^{DL-RL}} = y_s t_{f_s^{DL-RL}}. \quad (4.5)$$

**Si  $s$  est de type VBR :**

– sur les liens câblés :

$$\underline{\rho}_{f_s^{UL-WL}} y_s \leq x_{f_s^{UL-WL}} \leq y_s \bar{\rho}_{f_s^{UL-WL}}. \quad (4.6)$$

$$\underline{\rho}_{f_s^{DL-WL}} y_s \leq x_{f_s^{DL-WL}} \leq y_s \bar{\rho}_{f_s^{DL-WL}}. \quad (4.7)$$

– sur les liens radio :

$$\underline{t}_{f_s^{UL-RL}} y_s \leq x_{f_s^{UL-RL}} \leq y_s \bar{t}_{f_s^{UL-RL}}. \quad (4.8)$$

$$\underline{t}_{f_s^{DL-RL}} y_s \leq x_{f_s^{DL-RL}} \leq y_s \bar{t}_{f_s^{DL-RL}}. \quad (4.9)$$

Les contraintes qui testent les ressources pourraient forcer une variable  $x_f$  à zéro. Dans ce cas,  $y_s$  sera forcée à 0 et les  $x_f$  des autres flots aussi. La session sera ainsi refusée. Alors que si  $y_s = 1$ , on a  $x_f > 0$  pour les quatre flots de  $F_s$  et la session est acceptée.

Quand une session de type CBR est acceptée on a :

$$x_f = \rho_f \quad \text{pour } f \in F_s^{WL},$$

$$x_f = t_f \quad \text{pour } f \in F_s^{RL}.$$

Quand une session de type VBR est acceptée on a :

$$\underline{\rho}_f \leq x_f \leq \bar{\rho}_f \quad \text{pour } f \in F_s^{WL},$$

$$\underline{t}_f \leq x_f \leq \bar{t}_f \quad \text{pour } f \in F_s^{RL}.$$

L'ensemble de ces contraintes simulent donc le fonctionnement d'un contrôle d'admission : une session n'est acceptée que si les ressources sont suffisantes pour respecter les bornes

minimales de débit de ses quatre flots si elle est de type VBR ou les débits déclarés si elle est de type CBR. Si les ressources sont insuffisantes pour au moins un flot, alors la session est refusée et les débits de ses quatre flots sont forcés à 0.

#### 4.4.2 Niveaux de service

$$\frac{\sum_{s \in S^{a,p}} y_s}{|S^{a,p}|} \geq 1 - \text{Br}_{a,p} \quad a \in A, \quad p \in P. \quad (4.10)$$

Comme  $\text{Br}_{a,p}$  est le taux de blocage toléré,  $1 - \text{Br}_{a,p}$  représente le taux minimal d'acceptation des sessions. Cette contrainte force le taux réel des sessions acceptées d'application  $a$  et de priorité  $p$  à être supérieur ou égal au taux minimal.

#### 4.4.3 Liaison entre la partie câblée et la partie radio

Cette liaison se fait au niveau des stations de base et conditionne les débits  $x_f^i$  suivant les stations de base sélectionnées pour la session. Si la station de base  $i$  est sélectionnée, alors  $x_f^i$  doit prendre la valeur de débit déclarée sur les liens câblés quand la session est CBR ou une valeur continue entre deux bornes si la session est VBR. Ces bornes dépendent des RABs et ne sont pas les mêmes suivant le sens montant ou descendant.

**Si  $s$  est de type CBR :**

$$x_{f_s^{\text{UL-WL}}}^i = \rho_{f_s^{\text{UL-WL}}} \sum_{r \in T_s^{\text{UL}}} \alpha_{ir} f_s^{\text{UL-WL}} \quad s \in S^{\text{CBR}}, \quad i \in L_s^{\text{BS}}. \quad (4.11)$$

$$x_{f_s^{\text{DL-WL}}}^i = \rho_{f_s^{\text{DL-WL}}} \sum_{r \in T_s^{\text{DL}}} \alpha_{ir} f_s^{\text{DL-WL}} \quad s \in S^{\text{CBR}}, \quad i \in L_s^{\text{BS}}. \quad (4.12)$$

Si la station de base  $i$  est sélectionnée, alors  $\sum_{r \in T_s^{\text{UL}}} \alpha_{ir} f_s^{\text{UL-WL}} = 1$  et  $\sum_{r \in T_s^{\text{DL}}} \alpha_{ir} f_s^{\text{DL-WL}} = 1$ . Dans ce cas,  $x_f^i$  prend la valeur de débit déclaré  $\rho_f$ .

**Si  $s$  est de type VBR :**

– sens montant

$$\sum_{r \in T_s^{\text{UL}}} r_t \alpha_{ir} f_s^{\text{UL-WL}} \leq x_{f_s^{\text{UL-WL}}}^i \leq K_1 \sum_{r \in T_s^{\text{UL}}} r_t \alpha_{ir} f_s^{\text{UL-WL}} \quad s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}. \quad (4.13)$$

$K_1$  est une constante arbitrairement grande qui force  $x_{f_s^{UL-WL}}^i$  à zéro si la station de base  $i$  n'est pas sélectionnée pour la session  $s$ . Si elle est sélectionnée, le débit sur la partie câblée est supérieur ou égal à la valeur de débit du RAB choisi pour éviter toute accumulation au niveau de la station de base (formule (3.16)).

– sens descendant

$$0 \leq x_{f_s^{DL-WL}}^i \leq \sum_{r \in T_s^{DL}} r_t \alpha_{ir} f_s^{DL-RL} \quad s \in S^{VBR}, \quad i \in L_s^{BS}. \quad (4.14)$$

On a toujours  $x_{f_s^{DL-WL}}^i$  forcé à zéro si la station  $i$  n'est pas choisie. Mais dans ce sens, si la station de base  $i$  est sélectionnée, on veut que le débit du flot sur le lien câblé soit inférieur au débit du RAB choisi pour éviter à nouveau une accumulation à la station de base (formule (3.17)). C'est la contrainte sur le délai, décrite par la suite, qui va imposer une borne inférieure aux débits  $x_{f_s^{DL-WL}}^i$ .

#### 4.4.4 Qualité de service: délai

La contrainte sur le délai permet, via WFQ (formule (3.4)), d'imposer des débits minimaux aux flots d'une session sur les liens câblés. Cette contrainte s'applique donc aux sessions VBR pour lesquelles les débits peuvent prendre des valeurs continues. Dans le sens montant, le RAB est le seul critère sur lequel on peut jouer pour respecter la borne de délai. En effet, d'après la section 3.2.5, le délai de bout en bout pour un paquet se compose, dans ce sens, des délais de transmission, de propagation et de traitement sur les liens câblés et du délai sur le lien radio. Les délais de propagation et de traitement sont supposés constants pour un paquet. Le délai maximal de transmission est lié au débit minimal sur le lien câblé par la formule (3.23), il est donc aussi lié par la contrainte (4.13) au débit du RAB alloué. Le délai sur le lien radio est lié au FER du RAB par la formule (3.33). On garantit donc le délai dans le sens montant en forçant les RABs qui ont un débit suffisamment grand et un FER suffisamment petit pour respecter la borne de délai. Les sous-ensembles  $T_s^{UL}$  sont définis pour chaque session suivant le type de l'application tel que ces critères soient respectés. Les contraintes sur les choix des RABs pour chaque session s'appliquent à ces sous-ensembles. On garantit ainsi le délai dans le sens montant sans contrainte supplémentaire.

Dans le cas du sens descendant, la contrainte (4.14) n'impose pas de borne inférieure et le

débit du RAB alloué devient ici une borne supérieure. Les sous-ensembles  $T_s^{\text{DL}}$  sont définis tel que si le flot a un débit au plus égal à celui du RAB, la borne de délai soit respectée, sinon proposer ce RAB n'a aucun sens.

Il reste à définir une borne inférieure. Dans ce sens, le délai de bout en bout pour un paquet se compose du délai sur le lien radio, des délais de propagation et de transmission sur les liens câblés et des délais de traitement et d'attente accumulé sur tous les noeuds de sa route. Le délai d'attente est majoré par la borne de Parekh et Gallager donnée par la formule (3.30). On utilise cette borne pour minorer le débit sur les liens câblés.

Dans le pire cas nous avons :

$$\begin{aligned} D_{f_s^{\text{DL-WL}}}^{\text{trans},rt} &= D_{f_s^{\text{DL-WL}}}^{\text{max-trans},rt}, \\ D_{f_s^{\text{DL-WL}}}^{\text{attente}} &= D_s^*. \end{aligned}$$

D'après (3.24) et (3.4), on a :

$$D_{f_s^{\text{DL-WL}}}^{\text{max-trans},rt} = m_{f_s^{\text{DL}}} \times \frac{L_{f_s^{\text{DL-WL}}}}{\mathcal{R}(r_t^{\text{DL}})}, \quad (4.15)$$

$$D_{f_s^{\text{DL-WL}}}^{\text{attente}} = \frac{\sigma_s + m_{f_s^{\text{DL}}} L_{f_s^{\text{DL-WL}}}^{\text{max}}}{\rho_s}. \quad (4.16)$$

Initialement, dans la formule (3.4),  $K-1$  est le nombre de noeuds intermédiaires sur la route. Dans l'expression ci-dessus,  $m_{f_s^{\text{DL}}}$  est le nombre de liens de la route dans le sens descendant. Il y a donc  $m_{f_s^{\text{DL}}} - 1$  noeuds intermédiaires (le noeud Destination est exclu). Cependant, la route d'un flot ne comprend que les liens câblés du réseau de coeur (section 3.2.3). Il manque le lien câblé entre le BSC et la station de base. Or, il faut aussi comptabiliser le délai d'attente au niveau du BSC. Le nombre de noeuds considérés devient  $m_{f_s^{\text{DL}}}$ .

De plus, d'après la section 3.2.5, on a :

$$\rho_s = x_{f_s^{\text{DL-WL}}},$$

d'où :

$$D_{f_s^{\text{DL-WL}}}^{\text{attente}} = \frac{\sigma_s + m_{f_s^{\text{DL}}} L_{f_s^{\text{DL-WL}}}^{\text{max}}}{x_{f_s^{\text{DL-WL}}}}. \quad (4.17)$$

En combinant (3.20), (3.22) et (4.17) on obtient dans le pire cas :

$$D_s^{\text{DL}} = D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} + D_{f_s^{\text{DL-WL}}}^{\text{propag}} + D_{f_s^{\text{DL-WL}}}^{\text{trait}} + \frac{\sigma_s + m_{f_s^{\text{DL}}} L_{f_s^{\text{DL-WL}}}^{\text{max}}}{x_{f_s^{\text{DL-WL}}}} + D_{f_s^{\text{DL-RL}}}^r. \quad (4.18)$$

Si  $D_s^{\text{DL-MAX}}$  est le délai de transit maximal pour un paquet dans le sens descendant alors :

$$D_s^{\text{DL}} \leq D_s^{\text{DL-MAX}},$$

d'où :

$$x_{f_s^{\text{DL-WL}}} \geq \frac{m_{f_s^{\text{DL}}} L_{f_s^{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-RL}}}^r)}.$$

Les délais  $D_{f_s^{\text{DL-WL}}}^{\text{max-trans}}$  et  $D_{f_s^{\text{DL-RL}}}^r$  dépendent du RAB sélectionné (formules (3.24) et (3.34)).

La borne inférieure donnée par la contrainte ci-dessus dépend donc du choix du RAB. On choisit d'appliquer cette contrainte sur les variables  $x_{f_s^{\text{DL-WL}}}^i$ , en utilisant les variables  $\alpha_{ir}$  pour faciliter la dépendance de la borne au RAB. Par la suite, une contrainte permettra de relier le débit  $x_{f_s^{\text{DL-WL}}}^i$  au débit  $x_{f_s^{\text{DL-WL}}}$ .

La contrainte devient :

$$x_{f_s^{\text{DL-WL}}}^i \geq \sum_{r \in T_s^{\text{DL}}} \frac{m_{f_s^{\text{DL}}} L_{f_s^{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-RL}}}^r)} \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}. \quad (4.19)$$

A la section 3.2.2,  $\mathcal{R}(r_t)$  est aussi défini comme le débit minimal pour un flot  $f_s^{\text{DL-WL}}$  si la session est servie au débit  $r_t$  sur le lien radio.

Puisqu'il faut respecter simultanément ces deux bornes inférieures, la contrainte s'écrit finalement :

$$x_{f_s^{\text{DL-RL}}}^i \geq \sum_{r \in T_s^{\text{DL}}} \max \left( \mathcal{R}(r_t), \frac{m_{f_s^{\text{DL}}} L_{f_s^{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-RL}}}^r)} \right) \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}. \quad (4.20)$$

#### 4.4.5 Liaison entre le réseau de coeur et le réseau d'accès radio

Une fois le lien fait entre les débits  $x_{f_s^{\text{WL}}}^i$  et le RAB  $r$ , il faut relier les débits  $x_{f_s^{\text{WL}}}^i$  aux débits dans le réseau de coeur. Cette liaison se fait au niveau des BSCs.

La session, qu'elle soit CBR ou VBR, a le même débit sur tous les liens câblés de sa route jusqu'à la station de base. On a donc  $x_{f_s^{\text{WL}}} = x_{f_s^{\text{WL}}}^i$  avec  $i$  la station de base sélectionnée. La difficulté vient du fait, qu'en cas de *soft* handover, on peut avoir deux stations de base sélectionnées. La contrainte s'écrit alors intuitivement avec une forme quadratique :

$$(1 + \sum_{r \in T_s} z_{fr}) x_f = \sum_{j \in L_s^{\text{BS}}} x_f^j \quad f \in F_s^{\text{WL}}, \quad s \in S. \quad (4.21)$$

$T_s = T_s^{\text{UL}}$  ou  $T_s^{\text{DL}}$  suivant le sens du flot  $f$ .

En cas de *soft* handover, deux  $x_{f_s^{\text{WL}}}^i$  seront non nuls. Comme on minimise les capacités des liens câblés, on peut prédire que ces deux variables  $x_{f_s^{\text{WL}}}^i$  seront à leur valeur minimale (donnée par la contrainte (4.20)) et seront donc égales. On aura alors :  $x_{f_s^{\text{WL}}} = \frac{x_{f_s^{\text{WL}}}^{i_1} + x_{f_s^{\text{WL}}}^{i_2}}{2} = x_{f_s^{\text{WL}}}^{i_1}$  (ou  $x_{f_s^{\text{WL}}}^{i_2}$ ) avec  $i_1$  et  $i_2$  les deux stations de base sélectionnées.

On peut linéariser cette contrainte en écrivant :

$$x_f^i \leq x_f \leq \sum_{j \in L_s^{\text{BS}}} x_f^j \quad i \in L_s^{\text{BS}}, \quad f \in F_s^{\text{WL}}, \quad s \in S. \quad (4.22)$$

Parce que le modèle minimise les capacités des liens câblés, on s'attend au résultat suivant :

$$x_{f_s^{\text{WL}}} = \max_{(i \in L_s^{\text{BS}})} (x_{f_s^{\text{WL}}}^i).$$

Dans le cas du *soft* handover, on aura ainsi :

$$x_{f_s^{\text{WL}}} = x_{f_s^{\text{WL}}}^{i_1} \text{ (ou } x_{f_s^{\text{WL}}}^{i_2} \text{)}.$$

Les deux contraintes conduisent donc au même résultat.

La première étape pour la validation du modèle sera de s'assurer que le modèle se comporte conformément à nos attentes (section 5.4.1).

#### 4.4.6 Qualité de service: sélection du RAB sur les liens radio

$$\sum_{r \in T^{\text{UL}}} \alpha_{ir} f_s^{\text{UL-RL}} = \sum_{r \in T^{\text{DL}}} \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S, \quad i \in L_s^{\text{BS}}. \quad (4.23)$$

On s'assure avec cette contrainte que c'est la même station de base qui est sélectionnée pour le flot montant et pour le flot descendant d'une session acceptée.

$$x_f = \sum_{r \in T_s} r_t \left( \sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr} \right) \quad f \in F_s^{\text{RL}}, \quad s \in S. \quad (4.24)$$

$T_s = T_s^{\text{UL}}$  ou  $T_s^{\text{DL}}$  suivant le sens du flot  $f$ .

La variable  $x_f$  d'un flot RL prend ici une valeur discrète  $r_t$  correspondant au débit du RAB  $r$  sélectionné. On retranche  $z_{fr}$  car en *soft* handover au RAB  $r$  on a  $\sum_{i \in L_s^{\text{BS}}} \alpha_{irf} = 2$  et  $z_{fr} = 1$ . Si la session est refusée, toutes les variables sont à zéro et si la session est acceptée sans *soft* handover, on a  $\sum_{i \in L_s^{\text{BS}}} \alpha_{irf} = 1$  et  $z_{fr} = 0$ . En fait  $\sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr}$  est l'équivalent d'une variable binaire qui prend la valeur 1 si le RAB  $r$  est sélectionné qu'on soit en *soft* handover ou pas et 0 sinon.

$$\frac{\sum_{s \in S^{a,p}} \left( \sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr} \right) + (\text{Ordre}_{a,p}(r) - 1)}{\sum_{s \in S^{a,p}} y_s} \geq Q_{r,a,p} \quad f = f_s^{\text{UL-RL}} \text{ ou } f_s^{\text{DL-RL}}, \quad r \in T, \quad a \in A, \quad p \in P. \quad (4.25)$$

avec  $\sum_{s \in S^{a,p}} \left( \sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr} \right)$  le nombre de sessions acceptées d'application  $a$ , de priorité  $p$  et au RAB  $r$ . Pour chaque session, on considère indifféremment le flot  $f_s^{\text{UL-RL}}$  ou  $f_s^{\text{DL-RL}}$  pour les variables  $\alpha_{irf}$  et  $z_{fr}$ .  $\text{Ordre}_{a,p}(r)$  représente l'ordre de classement du RAB  $r$  dépendant du coefficient  $Q_{r,a,p}$  associé. Les RABs proposés pour chaque application sont classés par ordre décroissant suivant le coefficient  $Q_{r,a,p}$ . Ainsi, pour une application et une priorité donnée, le premier RAB est celui qui a le coefficient  $Q_{r,a,p}$  le plus élevé. Puis vient le second RAB, le troisième RAB et ce, jusqu'au dernier qui a le coefficient  $Q_{r,a,p}$  le plus petit.

Cette contrainte force à respecter les critères de qualité de service lors de l'attribution des RABs. En effet, on force la proportion des sessions acceptées au RAB  $r$  parmi les sessions acceptées d'application  $a$  et de priorité  $p$  à approcher le taux imposé  $Q_{r,a,p}$ . Pour le premier RAB, on a :  $\text{Ordre}_{a,p}(r) - 1 = 0$ , ainsi la contrainte force la proportion de sessions d'application  $a$  et de priorité  $p$  acceptées au RAB  $r$  à être supérieure ou égale au taux imposé  $Q_{r,a,p}$ . En revanche, pour les autres RABs,  $\text{Ordre}_{a,p}(r) - 1 > 0$ , ce qui implique que les taux effectifs peuvent être légèrement inférieurs au taux imposé  $Q_{r,a,p}$ .  $\text{Ordre}_{a,p}(r) - 1$

représente en fait le nombre de sessions en moins tolérées par rapport au taux voulu  $Q_{r,a,p}$ . Plus l'ordre du RAB est petit, plus le taux effectif peut être petit par rapport au taux imposé, privilégiant ainsi les RABs suivant leur ordre dans le classement.

#### 4.4.7 *soft* handover

$$\sum_{r \in T_s} z_{fr} \leq 1 \quad f \in F_s^{\text{RL}}, \quad s \in S. \quad (4.26)$$

$T_s = T_s^{\text{UL}}$  ou  $T_s^{\text{DL}}$  suivant le sens du flot  $f$ .

D'après la définition des variables  $z_{fr}$ , on a  $\sum_{r \in T_s} z_{fr} = 0$  quand la session est refusée ou acceptée sans *soft* handover. Quand  $\sum_{r \in T_s} z_{fr} = 1$ , la session est acceptée avec *soft* handover. Cette contrainte impose, dans ce cas, que pour le flot  $f$  sur un lien radio, il y ait au plus une variable  $z_{fr}$  à 1. On évite ainsi que plusieurs RABs soient choisis pour un même flot.

$$\sum_{i \in L_s^{\text{BS}}} \sum_{r \in T_s} \alpha_{irf} \leq 1 + \sum_{r \in T_s} z_{fr} \quad f \in F_s^{\text{RL}}, \quad s \in S. \quad (4.27)$$

Cette contrainte, combinée avec la contrainte (4.26), empêche que parmi l'ensemble des variables  $\alpha_{irf}$  d'une session  $s$ , plus de deux variables soient à 1 si la session est en *soft* handover. Cependant, cette contrainte ne garantit pas, dans ce cas, que les deux variables  $\alpha_{irf}$  seront au même RAB  $r$  et entre deux stations de base distinctes du sous-ensemble  $L_s^{\text{BS-SH}}$  de  $L_s^{\text{BS}}$ . Quand la session est acceptée et qu'il n'y a pas *soft* handover, la contrainte empêche que plus d'une variable  $\alpha_{irf}$  soit à 1. Avec les contraintes (4.4), (4.5), (4.8) et (4.9), on sait que si la session est acceptée alors les variables  $x_{f_s^{\text{UL-RL}}}$  et  $x_{f_s^{\text{DL-RL}}}$  sont strictement positives. De plus, grâce à la contrainte (4.24), on force au moins une variable  $\alpha_{itf}$  à 1. Ainsi, en cas de non *soft* handover, la contrainte ci-dessus garantit qu'une seule variable  $\alpha_{itf}$  sera à 1. On n'aura bien qu'une seule station de base sélectionnée et qu'un seul RAB pour le flot  $f$ .

$$z_{fr} = \frac{1}{2} \sum_{i \in L_s^{\text{BS-SH}}} w_{irf} \quad r \in T_s^{\text{UL}} \text{ ou } T_s^{\text{DL}}, \quad f \in F_s^{\text{RL}}, \quad s \in S. \quad (4.28)$$



Avec cette contrainte, quand  $z_{fr} = 1$ , on a exactement deux variables  $w_{irf}$  à 1 pour le flot  $f$ . Ces deux variables sont au même RAB  $r$  que  $z_{fr}$  et sont entre deux stations de base distinctes de la liste  $L_s^{\text{BS-SH}}$ . Dans le cas de *soft* handover, on a exactement une variable  $z_{fr}$  à 1 pour le flot  $f$  (contrainte (4.26)), on a alors exactement deux variables  $w_{irf}$  à 1 pour ce même flot  $f$ .

$$\alpha_{irf} \geq w_{irf} \quad r \in T, \quad i \in L_s^{\text{BS-SH}}, \quad f \in F_s^{\text{RL}}, \quad s \in S. \quad (4.29)$$

Cette contrainte, combinée avec les contraintes (4.27) et (4.28), force exactement deux variables  $\alpha_{irf}$  à 1 dans le cas de *soft* handover. Les variables  $\alpha_{irf}$  et  $w_{irf}$  forcées à 1 ont les même indices, ce qui est important pour les prochaines contraintes sur les capacités radio. On est assuré d'avoir deux variables  $\alpha_{irf}$  au même RAB et pour deux stations de base distinctes de  $L_s^{\text{BS-SH}}$ , ce que ne permettait pas de faire la contrainte (4.27).

On peut ici se poser la question de l'intérêt des variables  $w_{irf}$  qui sont fortement redondantes avec les variables  $\alpha_{irf}$ . En fait ces variables ont été introduites pour avoir des équations de capacité radio linéaires. La linéarité des contraintes a été prioritaire devant le nombre de variables et le nombre de contraintes.

#### 4.4.8 Capacités des liens câblés

Il faut faire ici la distinction entre les liens câblés du réseau de coeur et les liens câblés du réseau d'accès radio. Pour rappel, les liens câblés de la partie radio relient les BSC aux stations de base. Comme les stations de base sont choisies par l'algorithme, ces liens le sont aussi. Ils ne sont donc pas connus au départ. Ainsi, la route d'un flot  $f$  (paramètres de routage  $a_{ihf}$ ) inclut uniquement les liens câblés entre son noeud d'accès externe et le BSC, c'est-à-dire les liens du réseau de coeur (section 3.2.3). Pour dimensionner les liens câblés du réseau d'accès radio, il faut utiliser les variables  $x_f^i$  qui conditionnent la valeur du débit des flots  $f^{\text{WL}}$  sur ces liens suivant la station de base sélectionnée et les paramètres de routage  $b_{ihf}$  qui indiquent sur quelles périodes la session est en cours. Rappelons que les stations de base potentielles pour un usager mobile sont reliées à un même BSC. Si ce n'était pas le cas, la notion de lien potentiel s'appliquerait à des liens du réseau de coeur et il faudrait les traiter de la même façon que les liens câblés de la partie radio (section 3.5.2).

#### 4.4.8.1 Liens câblés du réseau de cœur

Pour un lien  $\ell$  du réseau de cœur :

$$\sum_{s \in S} \sum_{f \in F_s^{\text{WL}}} a_{\ell h f} x_f \leq C_\ell \quad h = 1, 2, \dots, H, \quad \ell = 1, 2, \dots, m. \quad (4.30)$$

La procédure d'optimisation minimise la capacité  $C_\ell$  telle que pour chaque période  $h$ , la somme des débits des flots qui traversent ce lien à la période  $h$  soit inférieure à la capacité  $C_\ell$ . Les liens sont supposés bidirectionnels, on somme donc indifféremment les débits des flots montant et descendant.

#### 4.4.8.2 Liens câblés du réseau d'accès radio

Pour chaque lien,  $\ell_{BS_i}$ , entre une station de base  $i$  et son BSC :

$$\sum_{s \in S} \sum_{f \in F_s^{\text{WL}}} b_{ih f} x_f^i \leq C_{\ell_{BS_i}} \quad h = 1, 2, \dots, H, \quad i = 1, 2, \dots, n_{\text{BS}}. \quad (4.31)$$

Dans le cas des liens câblés de la partie radio, on somme les variables  $x_f^i$  des flots. En effet, si la session est acceptée et que la station de base sélectionnée n'est pas  $i$  alors  $x_f^i = 0$ , la session ne contribuera donc pas au dimensionnement du lien correspondant. Par contre, si la station de base  $i$  est sélectionnée alors  $x_f^i = x_f$  et la session contribuera de la même façon au dimensionnement de ce lien que pour les autres liens du réseau de cœur.

#### 4.4.9 Capacités des liens radio

Les contraintes de capacités radio se basent sur les formules (3.7) et (3.13). Ces deux formules peuvent être interprétées de la façon suivante :

Le terme de droite représente la capacité radio maximale de la station de base dans le sens correspondant. Le terme de gauche qui apparaît dans la somme représente la capacité radio d'une session. On somme donc sur l'ensemble des sessions prises en charge par la station de base à une période donnée pour obtenir la capacité totale occupée. La contrainte impose que la capacité totale occupée par les sessions doit être inférieure à la capacité maximale de la station de base et ce pour chaque période. On a donc à quelques ajouts près nos contraintes de capacités radio.

Ces ajouts concernent le *soft* handover. On a vu que si la session est prise en *soft* handover par la station de base alors la capacité radio nécessaire est réduite (formules (3.9) et (3.15)). Les contraintes qui suivent permettent d'évaluer le bon montant de capacité radio d'une session selon qu'est en *soft* handover ou pas.

Les notations utilisées sont les mêmes que celles de la section 3.1.5.

#### 4.4.9.1 Capacité radio dans le sens montant

$$\frac{1+\lambda}{W} \sum_{s \in S} \sum_{r \in T_s^{\text{UL}}} \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}} \alpha_{irf_s^{\text{UL,RL}}}}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{UL}} \nu_s}{W}} - \frac{\gamma_{sr}^{\text{UL}} \nu_s (1 - SHG^{\text{UL}}) b_{ihf_s^{\text{UL,RL}}} w_{irf_s^{\text{UL,RL}}}}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} < L_{\text{UL}} \quad (4.32)$$

$i \in L_s^{\text{BS}}, \quad h = 1, 2, \dots, H.$

Si la session est refusée ou si elle n'est pas prise par la station de base  $i$  alors  $\alpha_{irf} = 0$  et  $w_{irf} = 0$ , elle n'est donc pas prise en compte dans le calcul de la capacité totale occupée pour la station de base  $i$ . Si la session est acceptée sans *soft* handover par la station de base  $i$  à un certain RAB  $r$ , alors  $\alpha_{irf} = 1$  et  $w_{irf} = 0$ . Le montant de la capacité radio occupée par cette session est alors :

$$\frac{1+\lambda}{W} \left( \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}} \times 1}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{UL}} \nu_s}{W}} - \frac{\gamma_{sr}^{\text{UL}} \nu_s (1 - SHG^{\text{UL}}) b_{ihf_s^{\text{UL,RL}}} \times 0}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} \right) = \frac{1+\lambda}{W} \left( \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}}}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{UL}} \nu_s}{W}} \right)$$

On retrouve la formule (3.8).

Maintenant, si la session est acceptée en *soft* handover alors  $\alpha_{irf} = 1$  et  $w_{irf} = 1$ . La capacité occupée par la session devient :

$$\begin{aligned}
& \frac{1+\lambda}{W} \left( \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}} \times 1}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{UL}} \nu_s}{W}} - \frac{\gamma_{sr}^{\text{UL}} \nu_s (1 - SHG^{\text{UL}}) b_{ihf_s^{\text{UL,RL}}} \times 1}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} \right) \\
&= \frac{1+\lambda}{W} \left( \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}} r_t (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W}) - \gamma_{sr}^{\text{UL}} \nu_s (1 - SHG^{\text{UL}}) b_{ihf_s^{\text{UL,RL}}}}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} \right) \\
&= \frac{1+\lambda}{W} \left( \frac{\gamma_{sr}^{\text{UL}} \nu_s SHG^{\text{UL}} (1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) b_{ihf_s^{\text{UL,RL}}}}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} \right) \\
&= \frac{1+\lambda}{W} \left( \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s b_{ihf_s^{\text{UL,RL}}}}{\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W}} \right)
\end{aligned}$$

On obtient bien la capacité attendue de la formule (3.9).

#### 4.4.9.2 Capacité radio dans le sens descendant

$$\begin{aligned}
& \sum_{s \in S} \sum_{r \in T_s^{\text{DL}}} \frac{\gamma_{sr}^{\text{DL}} \nu_s \left( N_0 a_{s,i} + \frac{P_{\text{BS}}}{W} \left( \sum_{j \in \text{Ring}(BS_0)} \frac{a_{s,i}}{a_{s,j}} + w \right) \right) b_{ihf_s^{\text{DL,RL}}} \alpha_{ir} f_s^{\text{DL,RL}}}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{DL}} \nu_s w}{W}} \\
& \frac{\gamma_{sr}^{\text{DL}} \nu_s \left( N_0 a_{s,i} + \frac{P_{\text{BS}}}{W} \left( \sum_{j \in \text{Ring}(BS_0)} \frac{a_{s,i}}{a_{s,j}} + w \right) \right) (1 - SHG^{\text{DL}}) b_{ihf_s^{\text{DL,RL}}} w_{ir} f_s^{\text{DL,RL}}}{(1 + \frac{\gamma_{sr}^{\text{DL}} \nu_s w r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{DL}} \gamma_{sr}^{\text{DL}} \nu_s w}{W})} < P_{\text{BS}} - P_{\text{cont}}
\end{aligned}$$

$$i \in L_s^{\text{BS}}, \quad h = 1, 2, \dots, H. \quad (4.33)$$

Le coefficient d'atténuation  $a_{s,i}$  est connu car les positions du terminal mobile  $m$  associé à la session  $s$  et de la station de base  $i$  sont fixées lors de la génération du trafic. Les positions des stations de base avoisinantes sont elles aussi connues mais on ne sait pas lesquelles seront sélectionnées. Les stations de base sont potentielles. Le calcul du rapport  $\sum_{j \in \text{Ring}(BS_0)} \frac{a_{s,i}}{a_{s,j}}$  est détaillé à la section 5.1.2.5. En suivant le même raisonnement que dans le sens montant, on obtient les capacités données par les formules (3.14) et (3.15) en remplaçant les valeurs des variables  $\alpha_{ir}$  et  $w_{ir}$  suivant si la session est en *soft* handover ou pas.

#### 4.4.10 Activation des stations de base potentielles

$$x^{BS_i} \leq \sum_{f \in F^{RL}} \sum_{r \in T} \alpha_{irf} \leq K_2 x^{BS_i} \quad i \in \mathcal{N}_{BS}. \quad (4.34)$$

Si une station de base potentielle a été sélectionnée au moins une fois pour une session quelconque, alors elle doit être activée ( $x^{BS_i} = 1$ ).  $x^{BS_i}$  deviendra une station de base à installer physiquement dans le dimensionnement optimal, si elle n'a jamais été sélectionnée alors  $x^{BS_i} = 0$ .

$K_2$  est une constante arbitrairement grande, au moins plus grande que deux fois le nombre total de sessions, qui est aussi le nombre maximal de flots possibles sur un lien radio sans tenir compte des critères de capacité.

### 4.5 Récapitulatif du modèle d'optimisation

#### 4.5.1 Définition et bornes des variables

- Variables associées aux flots sur les liens câblés du réseau de cœur :

$$x_f \in \mathbb{R}^+ \quad f \in F_s^{WL}, \quad s \in S.$$

- Variables associées aux flots sur les liens radio du réseau d'accès radio :

$$x_f \in \mathbb{R}^+ \quad f \in F_s^{RL}, \quad s \in S.$$

- Variables associées aux flots sur les liens câblés du réseau d'accès radio :

$$x_f^i \in \mathbb{R}^+ \quad f \in F_s^{WL}, \quad s \in S, \quad i \in L_s^{BS}.$$

- Variables pour le CAC :

$$y_s \in \{0,1\} \quad s \in S.$$

- Variables pour le choix des RABs et de la station de base :

$$\alpha_{irf} \in \{0,1\} \quad f \in F_s^{RL}, \quad s \in S, \quad i \in L_s^{BS}, \quad r \in T_s^{UL} \text{ ou } T_s^{DL}.$$

- Variables pour les capacités des liens câblés du réseau de cœur :

$$C_\ell \in \mathbb{R}^+ \quad \ell = 1, 2, \dots, m.$$

- Variables pour les capacités des liens câblés du réseau d'accès radio :

$$C_{\ell_{BS_i}} \in \mathbb{R}^+ \quad i \in \mathcal{N}_{BS}.$$

- Variables pour la sélection des stations de base :

$$x^{BS_i} \in \{0,1\} \quad i \in \mathcal{N}_{BS}.$$

– Variables pour le *soft* handover:

$$z_{fr} \in \{0,1\}, \quad f \in F_s^{\text{RL}}, \quad s \in S, \quad r \in T_s^{\text{UL}} \text{ ou } T_s^{\text{DL}}.$$

$$w_{irf} \in \{0,1\}, \quad f \in F_s^{\text{RL}}, \quad s \in S, \quad i \in L_s^{\text{BS-SH}}, \quad r \in T_s^{\text{UL}} \text{ ou } T_s^{\text{DL}}.$$

#### 4.5.2 Formulation mathématique du modèle

$$\min \left( w_c \sum_{\ell=1}^m d_\ell C_\ell + w_c \sum_{i=1}^{n_{\text{BS}}} d_{\ell_{BS_i}} C_{\ell_{BS_i}} + w_{\text{BS}} \sum_{i=1}^{n_{\text{BS}}} \mathcal{BS}_i \right)$$

##### CAC

$$\forall s \in S^{\text{CBR}} \text{ avec } F_s = \{f_s^{\text{UL-WL}}, f_s^{\text{DL-WL}}, f_s^{\text{UL-RL}}, f_s^{\text{DL-RL}}\}$$

$$(4.2) \quad x_{f_s^{\text{UL-WL}}} = y_s \rho_{f_s^{\text{UL-WL}}}$$

$$(4.3) \quad x_{f_s^{\text{DL-WL}}} = y_s \rho_{f_s^{\text{DL-WL}}}$$

$$(4.4) \quad x_{f_s^{\text{UL-RL}}} = y_s t_{f_s^{\text{UL-RL}}}$$

$$(4.5) \quad x_{f_s^{\text{DL-RL}}} = y_s t_{f_s^{\text{DL-RL}}}$$

$$\forall s \in S^{\text{VBR}} \text{ avec } F_s = \{f_s^{\text{UL-WL}}, f_s^{\text{DL-WL}}, f_s^{\text{UL-RL}}, f_s^{\text{DL-RL}}\}$$

$$(4.6) \quad \underline{\rho}_{f_s^{\text{UL-WL}}} y_s \leq x_{f_s^{\text{UL-WL}}} \leq y_s \bar{\rho}_{f_s^{\text{UL-WL}}}$$

$$(4.7) \quad \underline{\rho}_{f_s^{\text{DL-WL}}} y_s \leq x_{f_s^{\text{DL-WL}}} \leq y_s \bar{\rho}_{f_s^{\text{DL-WL}}}$$

$$(4.8) \quad \underline{t}_{f_s^{\text{UL-RL}}} y_s \leq x_{f_s^{\text{UL-RL}}} \leq y_s \bar{t}_{f_s^{\text{UL-RL}}}$$

$$(4.9) \quad \underline{t}_{f_s^{\text{DL-RL}}} y_s \leq x_{f_s^{\text{DL-RL}}} \leq y_s \bar{t}_{f_s^{\text{DL-RL}}}$$

##### Niveau de Service

$$(4.10) \quad \frac{\sum_{s \in S^{a,p}} y_s}{|S^{a,p}|} \geq 1 - \text{Br}_{a,p} \quad a \in A, \quad p \in P$$

### Soft handover

$$(4.26) \quad \sum_{r \in T_s} z_{fr} \leq 1 \quad f \in F_s^{\text{RL}}, \quad s \in S$$

$$(4.27) \quad \sum_{i \in L_s^{\text{BS}}} \sum_{r \in T_s} \alpha_{irf} \leq 1 + \sum_{r \in T_s} z_{fr} \quad f \in F_s^{\text{RL}}, \quad s \in S$$

$$(4.28) \quad z_{fr} = \frac{1}{2} \sum_{i \in L_s^{\text{BS-SH}}} w_{irf} \quad r \in T_s, \quad f \in F_s^{\text{RL}}, \quad s \in S$$

$$(4.29) \quad \alpha_{irf} \geq w_{irf} \quad r \in T_s, \quad i \in L_s^{\text{BS-SH}}, \quad f \in F_s^{\text{RL}}, \quad s \in S$$

### Capacités sur les liens radio

$$(4.23) \quad \sum_{r \in T_s^{\text{UL}}} \alpha_{ir} f_s^{\text{UL-RL}} = \sum_{r \in T_s^{\text{DL}}} \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S, \quad i \in L_s^{\text{BS}}$$

$$(4.24) \quad x_f = \sum_{r \in T_s} r_t \left( \sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr} \right) \quad f \in F_s^{\text{RL}}, \quad s \in S$$

$$(4.25) \quad \frac{\sum_{s \in S^{a,p}} \left( \sum_{i \in L_s^{\text{BS}}} \alpha_{irf} - z_{fr} \right) + (\text{Ordre}_{a,p}(r) - 1)}{\sum_{s \in S^{a,p}} y_s} \geq Q_{r,a,p}$$

$$f = f_s^{\text{UL-RL}} \text{ ou } f_s^{\text{DL-RL}}, \quad r \in T, \quad a \in A, \quad p \in P$$

$$(4.32) \quad \frac{1+\lambda}{W} \sum_{s \in S} \sum_{r \in T_s^{\text{UL}}} \frac{\gamma_{sr}^{\text{UL}} \nu_s b_{ih} f_s^{\text{UL-RL}} \alpha_{ir} f_s^{\text{UL-RL}}}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{UL}} \nu_s}{W}} - \frac{\gamma_{sr}^{\text{UL}} \nu_s (1 - SHG^{\text{UL}}) b_{ih} f_s^{\text{UL-RL}} w_{ir} f_s^{\text{UL-RL}}}{(1 + \frac{\gamma_{sr}^{\text{UL}} \nu_s r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{UL}} \gamma_{sr}^{\text{UL}} \nu_s}{W})} < L_{\text{UL}} \quad i \in \mathcal{N}_{\text{BS}}, \quad h \in H$$

$$(4.33) \quad \sum_{s \in S} \sum_{r \in T_s^{\text{DL}}} \frac{\gamma_{sr}^{\text{DL}} \nu_s \left( N_0 a_{s,i} + \frac{P_{\text{BS}}}{W} \left( \sum_{j \in \text{Ring}(BS_0)} \frac{a_{s,i}}{a_{s,j}} + w \right) \right) b_{ih} f_s^{\text{DL-RL}} \alpha_{ir} f_s^{\text{DL-RL}}}{\frac{1}{r_t} + \frac{\gamma_{sr}^{\text{DL}} \nu_s w}{W}} - \frac{\gamma_{sr}^{\text{DL}} \nu_s \left( N_0 a_{s,i} + \frac{P_{\text{BS}}}{W} \left( \sum_{j \in \text{Ring}(BS_0)} \frac{a_{s,i}}{a_{s,j}} + w \right) \right) (1 - SHG^{\text{DL}}) b_{ih} f_s^{\text{DL-RL}} w_{ir} f_s^{\text{DL-RL}}}{(1 + \frac{\gamma_{sr}^{\text{DL}} \nu_s w r_t}{W}) (\frac{1}{r_t} + \frac{SHG^{\text{DL}} \gamma_{sr}^{\text{DL}} \nu_s w}{W})} < P_{\text{BS}} - P_{\text{cont}} \quad i \in \mathcal{N}_{\text{BS}}, \quad h \in H$$

### Capacités sur les liens câblés

$$(4.22) \quad x_f^i \leq x_f \leq \sum_{j \in L_s^{\text{BS}}} x_f^j \quad s \in S, \quad f \in F_s^{\text{WL}}, \quad i \in L_s^{\text{BS}}$$

$$(4.11) \quad x_{f_s^{\text{UL-WL}}}^i = \rho_{f_s^{\text{UL-WL}}} \sum_{r \in T_s^{\text{UL}}} \alpha_{ir} f_s^{\text{UL-RL}} \quad s \in S^{\text{CBR}}, \quad i \in L_s^{\text{BS}}$$

$$(4.12) \quad x_{f_s^{\text{DL-WL}}}^i = \rho_{f_s^{\text{DL-WL}}} \sum_{r \in T_s^{\text{DL}}} \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S^{\text{CBR}}, \quad i \in L_s^{\text{BS}}$$

$$(4.13) \quad \sum_{r \in T_s^{\text{UL}}} r_t \alpha_{ir} f_s^{\text{UL-RL}} \leq x_{f_s^{\text{UL-WL}}}^i \leq M \sum_{r \in T_s^{\text{UL}}} r_t \alpha_{ir} f_s^{\text{UL-RL}} \quad s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}$$

$$(4.14) \quad x_{f_s^{\text{DL-WL}}}^i \leq \sum_{r \in T_s^{\text{DL}}} r_t \alpha_{ir} f_s^{\text{DL-RL}} \quad s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}$$

$$(4.20) \quad x_{f_s^{\text{DL-WL}}}^i \geq \sum_{r \in T_s^{\text{DL}}} \max \left( \mathcal{R}(r_t), \frac{m_{f_s^{\text{DL}}} L_{f_s^{\text{DL-WL}}}^{\max} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\max\text{-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-RL}}}^r)} \right) \alpha_{ir} f_s^{\text{DL-RL}} \\ s \in S^{\text{VBR}}, \quad i \in L_s^{\text{BS}}$$

$$(4.30) \quad \sum_{s \in S} \sum_{f \in F_s^{\text{WL}}} a_{\ell h f} x_f \leq C_\ell \quad h \in H, \quad \ell = 1, 2, \dots, m$$

$$(4.31) \quad \sum_{s \in S} \sum_{f \in F_s^{\text{WL}}} b_{ih f} x_f^i \leq C_{\ell_{BS_i}} \quad h \in H, \quad i = 1, 2, \dots, n_{\text{BS}}$$

### Sélection des stations de base

$$(4.34) \quad x^{BS_i} \leq \sum_{f \in F^{\text{RL}}} \sum_{r \in T} \alpha_{irf} \leq K x^{BS_i} \quad i \in \mathcal{N}_{\text{BS}}$$



## CHAPITRE 5

### RÉSULTATS

#### 5.1 Implémentation

L'implémentation du modèle d'optimisation, faite en C++, s'organise en deux parties distinctes:

- le générateur de trafic qui crée les jeux de données,
- le générateur de contraintes qui formule, à partir d'un jeu de données, les contraintes au format imposé par le logiciel d'optimisation CPLEX.

Le générateur de trafic crée un certain nombre de fichiers qui contiennent les informations sur les sessions. Ces fichiers servent ensuite de données d'entrée au générateur de contraintes. Le problème d'optimisation ainsi créé est alors résolu par CPLEX-MIP. Les sections suivantes décrivent en détail les fonctions et les paramètres liés à la création et à la résolution d'un problème.

##### 5.1.1 Générateur de trafic

Le générateur de trafic a trois fonctions:

- il fixe la durée de la planification ainsi que l'unité temporelle, c'est-à-dire la période élémentaire (section 3.2.4),
- il place les stations de base potentielles de façon uniforme sur la surface à couvrir,
- il crée une étude de cas pour le trafic en positionnant les usagers mobiles et en répartissant les sessions sur la durée de planification.

##### 5.1.1.1 Durée de la planification

La durée de la planification, évoquée à la section 3.2.4, est une donnée d'entrée du générateur de trafic. La durée de la période élémentaire doit aussi être fixée. Les arrivées et les durées des sessions seront des multiples de cette unité temporelle.

La durée de la période élémentaire détermine le nombre total de périodes sur la durée

de la planification. Ce nombre est critique car, d'un côté, plus il y a de périodes, plus les imprécisions sur le dimensionnement liées au séquençement temporel des sessions sont réduites (section 3.3.2). Mais d'un autre côté, le nombre de contraintes est aussi plus élevé. En effet, les contraintes (4.30) et (4.31) pour le dimensionnement des liens câblés et les contraintes (4.32) et (4.33) pour les capacités radio sont appliquées à chaque période élémentaire. Il faut donc choisir la durée de la période élémentaire en faisant un compromis entre la précision du séquençement temporel et la taille du problème.

Les deux paramètres d'entrée relatifs à cette fonction du générateur de trafic sont :

- DUREE\_PLANIFICATION,
- DUREE\_PERIODE.

#### 5.1.1.2 Carte géographique des stations de base potentielles

L'architecture du réseau de cœur ainsi que les BSCs sont fixés par la topologie du réseau. En revanche, le nombre de stations de base potentielles est une donnée d'entrée qui peut varier pour une même topologie de réseau. Les dimensions de l'aire géographique à couvrir sont aussi des données d'entrée. Le générateur de trafic répartit de façon uniforme les stations de base potentielles sur la surface à couvrir et calcule leurs positions. Avec une répartition homogène du trafic, on s'approche ainsi du cas idéal où les puissances émises par les stations de base sont identiques. Cette hypothèse, restrictive, est nécessaire aux formules de capacité radio décrites aux sections 3.1.5.3 et 3.1.5.4. On crée volontairement un sur-dimensionnement initial en choisissant un nombre élevé de stations de base potentielles (section 3.4).

Pour tous les résultats obtenus dans ce chapitre, on considère que l'aire géographique à couvrir ne contient pas de stations de base déjà existantes. Toutes les stations de base initiales sont donc potentielles. Le modèle pourrait cependant intégrer facilement le cas où des stations de base déjà installées seraient présentes.

Nous supposons que les BSCs se partagent l'aire géographique en couvrant des zones distinctes de surface à peu près équivalente. Ainsi, toutes les stations de base potentielles sont reliées au BSC qui couvre la zone où elles se trouvent.

Les paramètres d'entrée relatifs à cette fonction du générateur de trafic sont :

- LONGEUR\_SURFACE,
- LARGEUR\_SURFACE,

- NBR\_STATIONS\_POTENTIELLES.

### 5.1.1.3 Création des sessions

Le nombre de sessions est une donnée d'entrée au générateur de trafic. Plusieurs possibilités sont proposées :

- on fixe uniquement le nombre total de sessions et on sélectionne un scénario (A ou B),
- on fixe le nombre de sessions par type d'application.

Un scénario impose une certaine répartition des sessions par type d'application. De tels scénarios sont élaborés pour prévoir les profils des demandes et la nature des chargements de trafic pour la planification des réseaux. Ces scénarios se basent sur des études comportementales et des pronostics sur la façon dont les usagers utiliseront les applications. Les deux scénarios proposés dans le générateur de trafic s'inspirent de ceux proposés dans le rapport technique "*Application Traffic Model For UMTS Services*"<sup>[30]</sup>, rédigé par Ericsson et qui a pour but d'étudier les caractéristiques des nouvelles applications multimédia. Ce dernier décrit deux scénarios types où la contribution en termes de bande passante des quatre classes de trafic est évaluée en pourcentage du trafic total. Le premier scénario (A) prédit ainsi une forte utilisation des applications multimédia et semble s'appliquer particulièrement aux débuts de la 3G alors que le deuxième scénario (B) privilégie la voix tout en considérant une nette augmentation des applications multimédia par rapport à l'utilisation actuelle des services données. Ce dernier décrirait le trafic une fois que la 3G serait bien implantée dans notre vie quotidienne. Dans notre modèle, les quatre classes de trafic sont représentées par une application type (section 3.2.1), le profil de trafic se fait donc suivant ces applications. On va ainsi surestimer leurs contributions vu que dans le rapport technique<sup>[30]</sup>, une classe de trafic est un ensemble de plusieurs applications. En outre, la contribution en bande passante devient, dans notre modèle, une contribution en nombre d'utilisateurs en faisant la correspondance grâce aux débits moyens des applications. Le tableau 5.1 décrit la répartition des usagers par application pour les deux scénarios proposés par le générateur de trafic.

TAB. 5.1 – *Profils d'utilisation des applications par les usagers*

	Scenario A % des usagers	Scenario B % des usagers
<b>Voix</b>	40 %	60 %
<b>Vidéo conférence</b>	5 %	3 %
<b>Flux vidéo</b>	10 %	8 %
<b>Navigation Internet</b>	25 %	12 %
<b>Courrier électronique</b>	20 %	17 %

Les paramètres d'entrée relatifs à cette fonction du générateur de trafic sont :

- NBR\_TOTAL\_SESSIONS,
- SCENARIO.

ou

- NBR\_SESSIONS\_VOIX,
- NBR\_SESSIONS\_VIDEOCONF,
- NBR\_SESSIONS\_FLUX\_VIDEO,
- NBR\_SESSIONS\_NAVIGATION,
- NBR\_SESSIONS\_COURRIER.

Une fois le nombre de sessions par application fixé, le générateur de trafic crée les sessions.

Une session est la connexion entre deux entités, d'après la section 3.2.2, une session se compose de quatre ou deux flots suivant l'application. Les attributs d'une session sont résumés dans le tableau 5.2.

TAB. 5.2 – *Attributs d'une session suivant son application*

	Voix	Vidéo conférence	Flux vidéo	Navigation Internet	Courrier électronique
Début de la requête	oui	oui	oui	oui	oui
Durée de la connexion	oui	oui	oui	oui	oui
Priorité	oui	oui	oui	oui	oui
Type (High ou Low)	non	oui	oui	oui	non
Direction (UL ou DL)	non	non	oui (DL)	non	oui
Position	oui	oui	oui	oui	oui
Routes	oui	oui	oui	oui	oui

Pour donner des valeurs aux attributs d'une session, le générateur de trafic suit des règles détaillées dans les points suivants.

### Début des sessions

Le générateur calcule d'abord un ensemble d'intervalles entre les arrivées pour les requêtes

de connexion. Puis, il en déduit le séquençement temporel des sessions en faisant correspondre le début d'une requête au début d'une période. Deux distributions sont utilisées pour générer les temps d'inter-arrivée, on limite ainsi les symétries dans la génération du trafic.

– Distribution exponentielle.

Dans ce cas, les temps d'arrivée suivent une loi de Poisson. Si le taux d'arrivée des sessions est noté  $\lambda$ , alors le paramètre de la loi exponentielle est  $\lambda$ , signifiant que la moyenne des temps d'inter-arrivée est  $\frac{1}{\lambda}$  et la variance  $\frac{1}{\lambda^2}$ .

– Distribution lognormale.

Les paramètres de la loi lognormale sont la moyenne et la variance. On fixe ces deux paramètres aux mêmes valeurs que pour la loi exponentielle.

La résolution exacte du modèle est rendue difficile par le nombre élevé de variables binaires, ce qui a motivé le choix d'une période de planification réduite. Ainsi, le taux d'arrivée des sessions par type d'application est calculé simplement en appliquant  $\lambda = \frac{\text{NBR\_SESSIONS\_APPLICATION}}{\text{DUREE\_PLANIFICATION}}$ , fixant alors le taux d'arrivée à une valeur constante sur toute la durée de planification. La durée de planification réduite ne permet donc d'étudier qu'un seul profil de trafic et les variations observées sur une journée ne sont pas modélisées.

### Durée des sessions

On utilise une loi exponentielle dont le paramètre est  $\frac{1}{\text{DUREE\_MOY}}$ .

### Priorité, type et direction d'une session

Pour la priorité, le type et la direction d'une session, on utilise des lois de Bernoulli dont les paramètres représentent respectivement la probabilité de générer une session Gold, une session High et une session dont la direction est descendante (DL).  $(1 - \text{parametre})$  représente alors la probabilité d'avoir respectivement une session Silver, une session Low et une session dont la direction est montante (UL).

On rappelle que les sessions de flux vidéo sont uniquement dans le sens descendant et que le courrier électronique peut être dans les deux sens.

### Position d'une session

Le générateur de trafic répartit de façon uniforme sur la surface géographique les usagers mobiles faisant une requête pour la même application et à la même période . Ainsi, nous nous approchons d'une répartition homogène du trafic sur toutes les périodes. Une fois les coordonnées de l'utilisateur fixées, il faut déterminer l'ensemble des stations de base potentielles pour l'utilisateur mobile ( $L_s^{BS}$ ) ainsi que le sous-ensemble regroupant les stations de base potentielles qui pourront éventuellement prendre l'utilisateur en *soft* handover ( $L_s^{BS-SH}$ ). Ces stations de base potentielles doivent être reliées au même BSC pour éviter le multiroutage (section 3.5.2). Le choix des stations de base potentielles d'un usager se base sur un critère de distance. On définit deux distances de référence :

- $d_{BS-SH}$  définit le rayon maximal autour de l'utilisateur mobile, à l'intérieur duquel, les stations de base potentielles présentes sont jugées capables de prendre en charge la session avec ou sans *soft* handover.
- $d_{BS-NON-SH}$  définit le rayon du cercle autour du terminal mobile, à l'intérieur duquel, les stations de base potentielles présentes sont jugées capables de prendre en charge la session mais sans *soft* handover.

La figure 5.1 illustre le critère de sélection des stations de base potentielles pour un usager mobile.

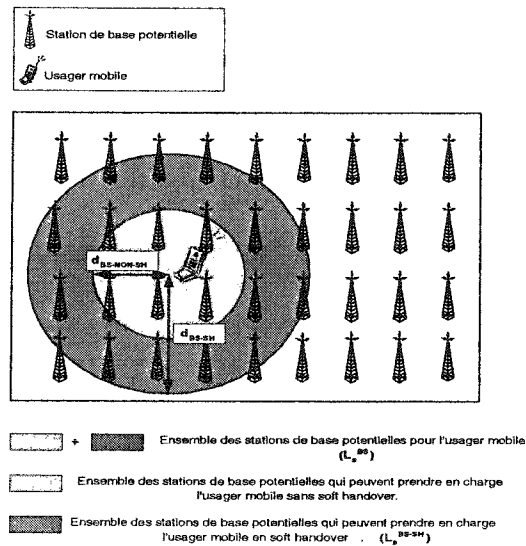


FIG. 5.1 – Sélection des stations de base potentielles pour un usager mobile

La distance  $d_{\text{BS-NON-SH}}$  est fixée à 30 mètres.

Suivant la surface géographique à couvrir et l'intensité du trafic, on peut estimer grossièrement le rayon moyen d'une cellule.  $d_{\text{BS-SH}}$  est alors fixé à deux fois ce rayon moyen. Ensuite, pour chaque station de base potentielle de l'ensemble  $L_s^{\text{BS}}$ , le coefficient d'atténuation du signal est évalué. Ce coefficient intervient dans les formules de capacité radio pour le sens descendant. D'après la formule (3.10), l'expression des coefficients d'atténuation est:

$$a_{s,i} = \frac{d^4(s,i)}{h_{\text{BS}i}^2} h_s^2$$

avec  $d(s,i)$  la distance entre l'utilisateur mobile et la station de base  $i$ ,  $h_{\text{BS}i}$  la hauteur de la station de base  $i$  et  $h_s$  la hauteur de l'utilisateur mobile.

On prend des valeurs moyennes pour les hauteurs :

$$h_s = 2 \text{ mètres},$$

$$h_{\text{BS}i} = 30 \text{ mètres}.$$

## Routes d'une session

Pour les routes d'une session, le générateur de trafic peut choisir une route différente pour les flots montant et descendant dans le cas d'une application interactive. Une station mobile est reliée au BSC qui couvre la zone où elle se trouve. Le générateur choisit alors aléatoirement pour chaque flot de la session une route entre ce BSC et un même nœud d'accès externe. L'annexe II détaille les valeurs des paramètres des lois de distribution utilisées dans le générateur de trafic pour créer les sessions.

### 5.1.2 Valeurs des paramètres dans les contraintes

#### 5.1.2.1 Valeurs des coûts dans la fonction objectif

La fonction objectif est donnée par la formule (4.1). Les coûts qui interviennent dans la fonction objectif sont liés au coût d'achat, au coût d'installation et de maintenance des stations de base et des liens câblés. L'influence de ces coûts sur le dimensionnement est étudiée à la section 5.4.2. Les valeurs des coefficients dans la fonction objectif sont :

$$- w_c = 1,$$

- $d_\ell = 1$ ,
- $w_{BS} = 500$ .

### 5.1.2.2 Bornes sur les débits des flots

- Les applications CBR.

TAB. 5.3 – Débits déclarés pour les flots des sessions CBR

	$\rho_{f_s^{UL-WL}}$ (kbps)	$\rho_{f_s^{DL-WL}}$ (kbps)	$t_{f_s^{UL-RL}}$ (kbps)	$t_{f_s^{DL-RL}}$ (kbps)
Voix	8	8	9.6	9.6
Vidéo conférence High	144	144	153.6	153.6
Vidéo conférence Low	64	64	76.8	76.8
Flux vidéo High	-	144	-	153.6
Flux vidéo Low	-	32	-	38.4

- Les applications VBR.

TAB. 5.4 – Bornes sur les débits pour les flots des sessions VBR

	$\underline{\rho}_{f_s^{UL-WL}} / \bar{\rho}_{f_s^{UL-WL}}$ (kbps)	$\underline{\rho}_{f_s^{DL-WL}} / \bar{\rho}_{f_s^{DL-WL}}$ (kbps)	$\underline{t}_{f_s^{UL-RL}} / \bar{t}_{f_s^{UL-RL}}$ (kbps)	$\underline{t}_{f_s^{DL-RL}} / \bar{t}_{f_s^{DL-RL}}$ (kbps)
Navigation Internet High	10/25	100/400	9.6/307.2	9.6/307.2
Navigation Internet Low	8/15	20/150	9.6/307.2	9.6/307.2
Courrier électronique	10/300	10 /300	9.6/153.6	9.6/153.6

### 5.1.2.3 Niveaux de service globaux

Les niveaux de service sont définis pour l'ensemble de la période de planification. Il s'agit d'un cas d'études avec des taux de rejet plus importants que ceux observés dans

TAB. 5.5 – Valeurs des taux de blocage  $Br_{ap}$

	Gold	Silver
Voix	10 %	30 %
Vidéo conférence	10 %	30 %
Flux vidéo	10 %	30 %
Navigation Internet	10 %	30 %
Courrier électronique	10 %	30 %

un réseau téléphonique (5% pour la voix), nous permettant d'étudier le comportement du contrôle d'admission avec des profils de trafic peu chargés.



### 5.1.2.4 Ensemble des RABs proposés par type d'application, valeurs des $\gamma_{sr}$ associés et des coefficients de qualité de service $Q_{r,a,p}$

Les valeurs cibles des rapports signal sur interférence définis pour chaque RAB  $r$ ,  $\gamma_{sr}$ , sont données sous forme d'abaques dans le rapport technique *IS-2000 Forward Link Simulation. Performance Results*<sup>[27]</sup> d'Ericsson. Les mesures sont faites pour un usager circulant à 3km/h (piéton) et un usager circulant à 30 km/h (voiture). Comme les usagers dans la partie radio sont supposés être tous statiques, nous avons retenu uniquement les mesures associées au piéton. Les usagers sont supposés statiques pour éviter le routage dynamique (section 3.2.3).

Les RABs proposés pour les applications de voix, de vidéo conférence et de flux vidéo dépendent des codecs utilisés. Pour les applications de navigation internet et de courrier électronique, les RABs proposés sont tels que les contraintes de délai soient réalistes (section 4.4.4).

TAB. 5.6 – Voix

	Voix				
	RAB $r$	$r_t$ (kbps); $r_{FER}$ (dB)	$\gamma_{sr}$	$Q_{r,a,p}$	
				Go	Si
$T_s^{UL}$	r1	[9.6; 1%]	1.7	90%	60%
	r2	[9.6; 2%]	1.2	10%	40%
$T_s^{DL}$	r1	[9.6; 1%]	3.7	90%	60%
	r2	[9.6; 2%]	3.2	10%	40%

TAB. 5.7 – Vidéo conférence

	Vidéo conférence Low					Vidéo conférence High				
	RAB $r$	$r_t$ (kbps); $r_{FER}$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$		RAB $r$	$r_t$ (kbps); $r_{FER}$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$	
				Go	Si				Go	Si
$T_s^{UL}$	r1	[76.8; 0.5%]	1.98	70%	30%	r1	[153.6; 0.5%]	2.11	70%	30%
	r2	[76.8; 1%]	1.78	30%	70%	r2	[153.6; 1%]	1.91	30%	70%
$T_s^{DL}$	r1	[76.8; 0.5%]	3.98	70%	30%	r1	[153.6; 0.5%]	4.11	70%	30%
	r2	[76.8; 1%]	3.78	30%	70%	r2	[153.6; 1%]	3.91	30%	70%

TAB. 5.8 – Flux vidéo

	Flux vidéo Low					Flux vidéo High				
	RAB $r$	$r_t$ (kbps); $r_{FER}$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$		RAB $r$	$r_t$ (kbps); $r_{FER}$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$	
				Go	Si				Go	Si
$T_s^{UL}$	r1	[38.4; 0.5%]	1.84	70%	30%	r1	[153.6; 0.5%]	2.11	70%	30%
	r2	[38.4; 1%]	1.57	30%	70%	r2	[153.6; 1%]	1.91	30%	70%
$T_s^{DL}$	r1	[38.4; 0.5%]	3.84	70%	30%	r1	[153.6; 0.5%]	4.11	70%	30%
	r2	[38.4; 1%]	3.57	30%	70%	r2	[153.6; 1%]	3.91	30%	70%

TAB. 5.9 – *Navigation Internet*

Navigation Internet Low						Navigation Internet High					
RAB r		$[r_t \text{ (kbps)}; \tau_{\text{FER}}]$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$		RAB r		$[r_t \text{ (kbps)}; \tau_{\text{FER}}]$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$	
				Go	Si					Go	Si
$T_s^{\text{UL}}$	r1	[9.6 ; 5%]	0.32	90%	10%	r1	[19.2 ; 5%]	0.35	90%	10%	
	r2	[9.6 ; 10%]	-0.81	10%	90%	r2	[19.2 ; 10%]	0.2	10%	90%	
$T_s^{\text{DL}}$	r1	[76.8 ; 5%]	3.12	60%	10%	r1	[307.2 ; 5%]	3.53	60%	10%	
	r2	[76.8 ; 10%]	2.81	25%	5%	r2	[307.2 ; 10%]	3.21	25%	5%	
	r3	[38.4 ; 5%]	2.95	10%	60%	r3	[153.6 ; 5%]	3.28	10%	60%	
	r4	[38.4 ; 10%]	2.61	5%	25%	r4	[153.6 ; 10%]	3.1	5%	25%	

TAB. 5.10 – *Courrier électronique*

Courrier électronique					
	RAB r	$[r_t \text{ (kbps)}; \tau_{\text{FER}}]$	$\gamma_{sr}$ (dB)	$Q_{r,a,p}$	
				Go	Si
$T_s^{\text{UL}}$	r1	[153.6; 5%]	1.28	40%	1.5%
	r2	[153.6; 10%]	1.1	25%	0.5%
	r3	[76.8; 5%]	1.12	15%	5%
	r4	[76.8; 10%]	0.81	10%	3%
	r5	[38.4; 5%]	0.95	5%	40%
	r6	[38.4; 10%]	0.61	3%	25%
	r7	[19.2; 5%]	0.88	1.5%	15%
	r8	[19.2; 10%]	0.48	0.5%	10%
$T_s^{\text{DL}}$	r1	[153.6; 5%]	1.28	40%	1.5%
	r2	[153.6; 10%]	3.1	25%	0.5%
	r3	[76.8; 5%]	3.12	15%	5%
	r4	[76.8; 10%]	2.81	10%	3%
	r5	[38.4; 5%]	2.95	5%	40%
	r6	[38.4; 10%]	2.61	3%	25%
	r7	[19.2; 5%]	2.88	1.5%	15%
	r8	[19.2; 10%]	0.48	0.5%	10%

#### 5.1.2.5 Valeurs des constantes dans les formules de capacités radio

Les valeurs numériques des constantes qui apparaissent dans les formules de capacité radio proviennent de *WCDMA for UMTS*<sup>[4]</sup>.

- $W=1.25$  MHz (bande passante de CDMA2000)
- $N_0=1.2589 \cdot 10^{-20}$  W/Hz
- $\lambda = 55$  %
- $L_{\text{UL}} = 60$  %
- $P_{\text{BS}} = 20$  Watts
- $P_{\text{cont}} = 20$  %  $P_{\text{BS}} = 4$  Watts
- $w = 0.4$
- $SHG^{\text{UL}} = 45$  % correspondant à un gain de 3.5 dB (section 3.1.4).
- $SHG^{\text{DL}} = 45$  %

- $\nu_s = 0.67$  pour les sessions de voix et  $\nu_s = 1$  pour toutes les autres applications.

Dans la contrainte (4.33) pour les capacités radio dans le sens descendant, le coefficient d'atténuation  $a_{s,i}$  est une valeur connue, calculée dans le générateur de trafic à partir des positions de l'utilisateur mobile et de la station de base potentielle. En revanche, la valeur exacte du rapport  $\sum_j \frac{a_{s,i}}{a_{s,j}}$  n'est pas connue à l'avance car les stations de base voisines  $j$  ne sont pas encore déterminées. Pour contourner cette difficulté, nous utilisons le résultat obtenu dans [28] qui calcule le rapport moyen  $E[\frac{a_{s,i}}{a_{s,j}}]$  en considérant des cellules hexagonales et en tenant compte des 6 premières cellules avoisinantes. La valeur obtenue est  $E[\frac{a_{s,i}}{a_{s,j}}] = 0.117$ . Nous utilisons alors cette valeur moyenne pour obtenir une valeur approchée des coefficients d'atténuation entre l'utilisateur mobile et les stations de base voisines. Nous faisons l'approximation suivante :  $a_{s,j} = \frac{a_0}{0.117}$  avec  $a_0$  la valeur du coefficient d'atténuation pour un usager mobile éloigné de sa station de base mère d'une distance égale à la moitié du rayon moyen d'une cellule, estimé à priori compte tenu de la surface à couvrir et de l'intensité du trafic. Ce rayon moyen est aussi utilisé dans le générateur de trafic pour construire l'ensemble  $L_s^{\text{BS}}$  (section 5.1.1.3).

$\sum_j \frac{a_{s,i}}{a_{s,j}}$  devient alors, en considérant 6 cellules avoisinantes,  $6 a_{s,i} \times \frac{0.117}{a_0}$ .

Il aurait été plus simple d'approcher directement le rapport  $\sum_j \frac{a_{s,i}}{a_{s,j}}$  par  $6 E[\frac{a_{s,i}}{a_{s,j}}]$ . L'intérêt de notre approximation est de conserver une relation entre la capacité radio d'une session et la position de l'utilisateur mobile en conservant le coefficient  $a_{s,i}$ . Dans la contrainte (4.33), le coefficient  $a_{s,i}$  multiplié par  $N_o$  est rendu négligeable, la dépendance de cette formule à la position des usagers mobiles est donc limitée à la somme  $\sum_j \frac{a_{s,i}}{a_{s,j}}$ . D'où l'intérêt de la conserver par le biais de notre approximation. La dépendance par rapport aux positions des usagers mobiles est importante car sans cela, la procédure choisirait n'importe quelle station de base pour assurer la connexion.

#### 5.1.2.6 Valeurs des paramètres pour la contrainte du délai

La contrainte du délai (4.20) s'applique aux flots descendant des applications VBR, c'est-à-dire, dans notre modèle, aux applications de navigation Internet et aux courriers électroniques.

Les tableaux 5.11 et 5.12 donnent, pour ces deux applications, les valeurs numériques des

paramètres qui interviennent dans les calculs des délais décrits dans les sections 3.2.5 et 4.4.4.

TAB. 5.11 – Paramètres pour le délai d'une session de navigation Internet

	Navigation Internet Low	Navigation Internet High
$L_{f_{DL},WL}^{max}$	1460 octets	1460 octets
$L_{f_{DL},WL}$	800 octets	800 octets
$\sigma_s$	1460 octets	1460 octets
Temps_propagation/liens	20 ms/noeud	20 ms/noeud
Temps_traitement/routeur	10 ms/liens	10 ms/liens
Délai_max/page	6 s/page	15 s/page
Nbr_paquets_moyen/page	3 paquets/page	3 paquets/page
$D_s^{DL-MAX} = \frac{\text{Délai\_max/page}}{\text{Nbr\_paquets\_moyen/page}}$	5 s	2 s

TAB. 5.12 – Paramètres pour le délai d'une session de courrier électronique

	Courrier électronique
$L_{f_{DL},WL}^{max}$	1460 octets
$L_{f_{DL},WL}$	800 octets
$\sigma_s$	1460 octets
Temps_propagation/liens	20 ms/noeud
Temps_traitement/routeur	10 ms/liens
$D_s^{DL-MAX}$	10 s

L'évaluation des délais dépend, entre autres, du nombre de liens câblés sur la route du flot. L'annexe III donne un exemple de calcul des coefficients de la contrainte (4.20).

### 5.1.3 Procédure d'optimisation : CPLEX

#### Généralités

CPLEX est un logiciel d'optimisation développé par ILOG<sup>[31]</sup>. CPLEX peut résoudre des problèmes d'optimisation linéaires et mixtes (PLM) grâce à sa version MIP (*Mixed Integer Programming*) que nous appliquons ici à un problème mixte 0-1. Le manuel de référence <sup>[32]</sup> explique en détail les fonctions des paramètres qui permettent aux usagers de personnaliser leur utilisation du logiciel. Par la suite, nous expliquons succinctement le fonctionnement de CPLEX-MIP ainsi que les paramètres qui seront utilisés dans le cadre de notre étude. CPLEX-MIP utilise un algorithme de séparation et d'évaluation qui résout une série de sous-problèmes linéaires en nombres entiers. A l'étape initiale et à chaque itération, il procède en résolvant la relaxation continue du problème ou sous-problème. La relaxation continue consiste à résoudre le problème après avoir relaxé les variables binaires en variables conti-

nues comprises entre 0 et 1. Pour un problème de minimisation, la solution optimale de la relaxation continue représente une borne inférieure pour la solution optimale en nombres entiers. Une solution en nombres entiers ou solution entière est une solution pour laquelle toutes les variables binaires, initialement relaxées, sont à 0 ou à 1.

La séparation consiste ensuite à diviser le problème en plusieurs sous-problèmes dans lesquels certaines variables binaires seront fixées à 0 ou à 1. Les solutions de ces sous-problèmes permettent d'encadrer la solution optimale en nombres entiers. Chaque sous-problème est évalué par relaxation continue. En réduisant l'espace des solutions, les sous-problèmes permettent d'augmenter la borne inférieure pour mieux borner la solution entière optimale. De plus, chaque solution entière trouvée permet possiblement d'améliorer la valeur de la borne supérieure.

### **Solution $\epsilon$ -optimale**

La difficulté est d'améliorer la borne inférieure en se plaçant dans le bon espace de solutions. Pour gérer l'ensemble des sous-problèmes, CPLEX-MIP construit, au fur et à mesure, un arbre de branchement où chaque nœud est un sous-problème avec une nouvelle variable binaire fixée. La racine de l'arbre est le problème initial. Chaque relaxation continue à un nœud fournit une borne inférieure pour tout le sous-arbre. A la première solution entière trouvée, la différence relative entre l'objectif de cette solution (qui devient la borne supérieure) et la plus petite borne inférieure parmi les nœuds encore inexplorés est conservée. Cette différence relative ou écart est réévaluée à chaque nouveau branchement ou quand une solution entière est trouvée. La procédure s'arrête quand cet écart est inférieur à une certaine valeur que l'on fixe (paramètre `CPX_PARAM_EPAGAP`). La procédure rend alors comme solution finale la meilleure solution entière trouvée. Si on fixe l'écart à zéro, on est assuré d'avoir la solution optimale car on sait qu'il n'existe pas de meilleure solution dans les branches non explorées. Si l'écart est strictement positif, on a alors une solution  $\epsilon$ -optimale dont la précision  $\epsilon$  est l'écart toléré.

### **Elimination implicite des sous-problèmes**

A chaque nœud de l'arbre, CPLEX-MIP résout le sous-problème par relaxation continue. Pour chaque sous-problème, la solution trouvée est soit entière, soit fractionnaire ou il n'y a pas de solution. Si la solution est entière, la borne supérieure est actualisée et la procédure continue sur une autre branche. Un sous-arbre peut être abandonné si la relaxation continue

au nœud parent fournit une solution plus grande que la borne supérieure. En effet, la relaxation continue fournit une borne inférieure pour tout le sous-arbre, on sait alors qu'il ne contiendra pas de meilleure solution. Si la solution est fractionnaire, alors la procédure peut soit continuer le branchement plus profondément dans la branche soit remonter dans l'arbre. Dans le cas où elle continue dans la branche, CPLEX-MIP choisit une variable fractionnaire et crée deux nouveaux sous-problèmes en forçant la variable à 0 et à 1. Avant la relaxation continue, la procédure d'optimisation cherche des coupes pour réduire le domaine réalisable. Une coupe est une contrainte supplémentaire que CPLEX-MIP ajoute pour exprimer certaines relations entre les variables binaires, par exemple quand, dans un sous-ensemble de variables binaires, il peut y avoir au plus une variable à 1. Les différentes coupes possibles sont décrites en détails dans [32]. Il est possible de forcer CPLEX-MIP à générer plus ou moins de coupes pendant le parcours de l'arbre (paramètres `CPX.PARAM.CLQUES`, `CPX.PARAM.COVERS` et `CPX.PARAM.GUBCOVERS`). Les coupes permettent de réduire le nombre de variables fractionnaires de la relaxation continue et limitent les prochaines possibilités de branchement.

Il est possible, avec le paramètre `CPX.PARAM.CUTUP`, de fixer une valeur d'abandon qui sert de borne supérieure pour l'objectif de la relaxation continue à tous les nœuds de l'arbre. Si l'objectif est supérieur à la valeur de l'abandon, alors la procédure rejette le nœud et remonte dans l'arbre. La valeur de l'abandon est par défaut très élevée, mais si par la nature du problème on connaît une limite physique à la valeur de l'objectif optimal, alors cette stratégie peut être intéressante car elle réduit les branches et permet donc d'obtenir la solution optimale plus rapidement. En revanche, si la valeur d'abandon est trop restrictive, alors la procédure pourrait passer à côté de la solution optimale.

### **Stratégies d'exploration, de séparation et de branchement**

Il est possible, grâce à certains paramètres, de contrôler la construction de l'arbre de branchement ainsi que le parcours de l'arbre. Avec ces paramètres, on peut choisir plusieurs stratégies dans les choix qui se posent.

- Dans l'arbre, quel nœud doit-on parcourir en premier? (paramètres `CPX.PARAM.BTTOL` et `CPX.PARAM.NODESEL`)
- Au niveau du nœud, sur quelle variable doit-on séparer pour construire les nœuds fils? (paramètre `CPX.PARAM.VARSEL`)

- Au niveau de la variable, quel nœud fils doit être créé en premier : variable à 0 ou à 1? (paramètre CPX.PARAM.BRDIR)

Les différentes stratégies proposées pour chaque paramètre sont :

- **CPX.PARAM.BTTOL**

Ce paramètre contrôle le retour en arrière dans le parcours de l'arbre. Une valeur faible tend à favoriser le retour en arrière alors qu'une valeur élevée favorisera la recherche en profondeur. Par défaut, la valeur de ce paramètre est faible.

- **CPX.PARAM.NODESEL**

Ce paramètre agit sur le choix du prochain nœud déjà construit mais inexploré. Quatre stratégies d'exploration sont proposées :

- **La recherche en profondeur.** Dans ce cas, la procédure choisit le nœud le plus récent dans la construction de l'arbre, elle explore donc les branches le plus profond possible.
- **La recherche suivant les meilleures bornes.** Dans ce cas, la procédure choisit le nœud avec la meilleure borne inférieure obtenue par relaxation continue.
- **La recherche suivant la meilleure estimation de la solution.** Dans ce cas, la procédure choisit le nœud pour lequel l'estimation d'une solution entière est la meilleure.
- **La recherche alternative suivant la meilleure estimation de la solution.** Dans ce cas, la procédure choisit le nœud pour lequel l'estimation d'une solution entière est la meilleure une fois que toutes les combinaisons entières non réalisables sont supprimées de l'arbre.

Par défaut, CPLEX-MIP utilise la recherche suivant les meilleures bornes.

- **CPX.PARAM.VARSEL**

Ce paramètre contrôle quelle variable choisir lorsque un nouveau nœud doit être créé.

Les stratégies de séparation retenues dans notre étude sont :

- **Séparation suivant la variable de valeur fractionnaire maximale,** c'est-à-dire sur la variable qui a la valeur fractionnaire la plus élevée (la plus proche de 0.5).
- **Séparation suivant la variable de valeur fractionnaire minimale,** c'est-

à-dire sur la variable qui a la valeur fractionnaire la plus proche de 0 ou 1.

- **Séparation prédictive.** Ce branchement se base sur des estimations pour savoir quel nœud fils créé est le plus prometteur.
- **Séparation automatique.** La procédure évalue la meilleure stratégie et s'adapte en fonction.

Par défaut, CPLEX-MIP utilise la séparation automatique.

#### – CPX\_PARAM\_BDDIR

Ce paramètre contrôle le nœud fils qui sera créé en premier. CPLEX-MIP propose trois stratégies de branchement :

- **SUP:** la variable est d'abord fixée à 1.
- **INF:** la variable est d'abord fixée à 0.
- **automatique:** CPLEX évalue la meilleure valeur et choisit en fonction.

Par défaut, CPLEX-MIP choisit de façon automatique.

En plus du paramètre CPX\_PARAM\_VARSEL, CPLEX-MIP permet de mettre des priorités sur les variables. Ainsi, la procédure va privilégier dans la séparation les variables fractionnaires qui ont une forte priorité. Ce système de priorité s'associe aux stratégies du paramètre CPX\_PARAM\_VARSEL quand la procédure doit choisir parmi plusieurs variables qui ont la même priorité.

Suivant la forme du problème mathématique, les stratégies sont plus ou moins efficaces. Dans la section 5.3, nous cherchons la combinaison des stratégies de CPLEX-MIP la plus adaptée à notre problème.

## 5.2 Description des jeux de données

Un jeu de données est défini par les trois critères suivants:

- la topologie du réseau de cœur et des BSCs,
- le profil de trafic,
- le nombre de stations de base potentielles.



### 5.2.1 Topologies tests des réseaux de cœur et des BSCs

La topologie des réseaux de cœur et des BSCs définit l'agencement des liens câblés des nœuds d'accès externe jusqu'aux BSCs. La topologie ne contient pas les liens câblés entre les BSCs et les stations de base car ces liens n'existent que si la station de base correspondante est sélectionnée. Pour la plupart des études de cas, nous utilisons la topologie test réseau01 décrite ci-dessous. Une topologie plus importante, réseau02, est utilisée à la dernière section 5.5.3 pour étudier l'évolution du dimensionnement en fonction de la charge et du profil de trafic.

#### – Réseau01

- réseau de cœur : 2 PDSNs, 3 nœuds intermédiaires, 8 liens câblés,
- réseau d'accès radio : 2 BSCs, 2 liens câblés,
- nombre de liens sur les routes : 4 (incluant le lien entre le BSC et la station de base).

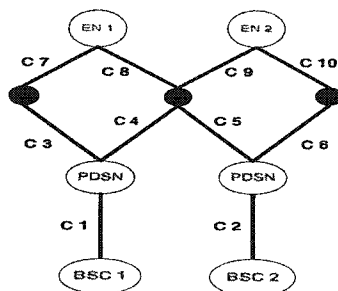


FIG. 5.2 – Exemple de réseau réseau01

#### – Réseau02

- réseau de cœur : 2 PDSNs, 7 nœuds intermédiaires, 18 liens câblés,
- réseau d'accès radio : 2 BSCs, 2 liens câblés,
- nombre de liens sur les routes : 5 (incluant le lien entre le BSC et la station de base).

### 5.2.2 Profil de trafic

Le profil de trafic définit la répartition des sessions selon les cinq applications proposées. De plus, il fixe toutes les informations liées aux requêtes : le type d'application, les ensembles

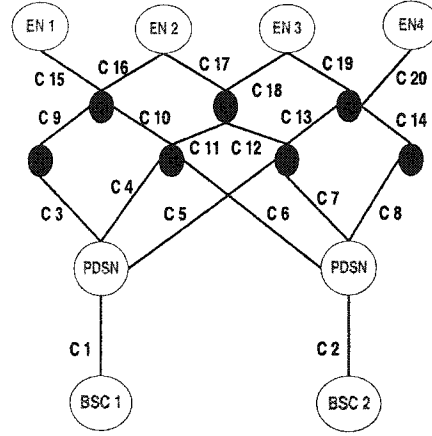


FIG. 5.3 – Exemple de réseau réseau02

des RABs disponibles ( $T_s^{UL}$  et  $T_s^{DL}$ ), la priorité de l'utilisateur, la direction pour le courrier électronique, le début et la durée de la connexion et la position de l'utilisateur.

Pour certains profils de trafic, la répartition des sessions Low et High pour les applications de vidéo conférence, de flux vidéo et de navigation Internet est désactivée. Ainsi, pour ces applications, il n'y a que des sessions High ou Low. De plus, pour réduire le nombre de variables binaires, certains jeux de données se limitent à un ou deux RABs les ensembles  $T_s^{UL}$  et  $T_s^{DL}$  décrits dans les tableaux 5.6, 5.7, 5.8, 5.9 et 5.10. Les coefficients de qualité de service donnés dans ces tableaux ne sont alors plus les mêmes.

- Pour les applications de navigation Internet et de courrier électronique, quand les ensembles  $T_s^{UL}$  et  $T_s^{DL}$  sont réduits à deux RABs, les deux coefficients  $Q_{r,a,p}$  sont 90 % pour le RAB impliquant la meilleure qualité de service et 10 % pour le second.
- Quand les ensembles  $T_s^{UL}$  et  $T_s^{DL}$  sont réduits à un seul RAB, la contrainte de qualité de service est désactivée pour l'application concernée.

### 5.2.3 Nombre de stations de base potentielles

Le nombre de stations de base potentielles agit sur les positions discrètes de ces dernières et sur les ensembles  $L_s^{BS}$  et  $L_s^{BS-SH}$  définis pour chaque usager mobile. En effet, quand on change un jeu de données en modifiant le nombre de stations de base potentielles, ces dernières sont redistribuées de façon uniforme sur l'aire géographique. Leurs positions sont alors changées, ce qui entraîne des changements pour les ensembles  $L_s^{BS}$  et  $L_s^{BS-SH}$  de chaque

usager mobile. De plus, le nombre total de liens câblés change aussi car, en plus des liens câblés fixés par la topologie du réseau, il faut compter les liens câblés entre les stations de base potentielles et leurs BSCs attitrés.

#### 5.2.4 Description des jeux de données

Les tableaux 5.13 et 5.14 détaillent les différents paramètres pour des jeux de données résolus dont les résultats sont analysés dans les sections suivantes. En particulier, pour chaque instance, les RABs proposés par application sont précisés.

TAB. 5.13 – *Instances (01)1-8, (01)1-30, (01)2-16*

	Instance (01)1-8			Instance (01)1-30			Instance (01)2-16		
Topologie réseau	réseau01			réseau01			réseau01		
Aire géographique	500 m × 400 m			500 m × 400 m			500 m × 400 m		
Stations de base potentielles	8			30			16		
Liens câblés	18			40			26		
Durée de la planification	1 minute			1 minute			1 minute		
Durée d'une période	10 secondes			10 secondes			10 secondes		
Nombre de sessions	60	RAB		60	RAB		80	RAB	
		$r_t$ (kbps)	$r_{FER}$ (%)		$r_t$ (kbps)	$r_{FER}$ (%)		$r_t$ (kbps)	$r_{FER}$ (%)
Voix	30	9.6	1	30	9.6	1	80	9.6	1, 2
Vidéo conférence Low	3	76.8	0.5	3	76.8	0.5	0	-	-
Flux vidéo Low	7	38.4	0.5	7	38.4	0.5	0	-	-
Navigation Internet High	12	UL 19.2	5	12	UL 19.2	5	0	-	-
		DL 153.6	5		DL 153.6	5		-	-
Courrier électronique	8	76.8	5	8	76.8	5	0	-	-
# variables	1554			4793			6522		
# variables binaires	906			2971			4896		
# contraintes	2435			6564			6950		

Le système de notation des jeux de données consiste à nommer une instance (01)1-8, par exemple, l'instance qui se base sur la topologie réseau01, sur le profil de trafic 1 avec 8 stations de base potentielles. Ainsi, les instances (01)1-8 et (01)1-30 se basent sur la même topologie de réseau et sur le même profil de trafic 1. Ceci signifie que la répartition des sessions par application ainsi que toutes les caractéristiques liées aux requêtes sont identiques pour ces deux instances. Seul le nombre de stations de base potentielles change. Notons que les instances (01)2-16 et (01)3-16 ont deux RABs par application (même débit mais FER différent).

A chaque profil de trafic correspond une certaine répartition des sessions par application mais aussi un séquençement temporel des sessions. Ce séquençement consiste à faire correspondre les débuts et les durées des sessions avec des multiples de périodes élémentaires

TAB. 5.14 – Instances (01)3-16 , (01)4-20 et (02)5-25

	Instance (01)3-16			Instance (01)4-20			Instance (02)5-25		
Topologie réseau	réseau01			réseau01			réseau02		
Aire géographique	500 m×400 m			500 m×400 m			500 m×400 m		
Stations de base potentielles	16			20			25		
Liens câblés	26			30			45		
Durée de la planification	1 minute			1 minute			1 minute		
Durée d'une période	10 secondes			10 secondes			10 secondes		
Nombre de sessions	80	RAB		150	RAB		150	RAB	
		$r_t$ (kbps)	$r_{FER}(\%)$		$r_t$ (kbps)	$r_{FER}(\%)$		$r_t$ (kbps)	$r_{FER}(\%)$
Voix	30	9.6	1, 2	60	9.6	1	89	9.6	1
Vidéo conférence Low	7	76.8	0.5, 1	15	76.8	0.5	3	76.8	0.5
Vidéo conférence High	0	-	-	0	-	-	4	153.6	0.5
Flux vidéo Low	10	38.4	0.5, 1	20	38.4	0.5	7	38.4	0.5
Flux vidéo High	0	-	-	0	-	-	7	153.6	0.5
Navigation Internet Low	0	-	-	0	-	-	5	UL 9.6	5
Navigation Internet High	23	UL 19.2	5, 10	30	UL 19.2	5	10	DL 76.8	5
Courrier électronique	10	DL 307.2	5, 10	25	DL 153.6	5	25	UL 19.2	5
		153.6	5, 10		76.8	5		DL 307.2	5
# variables	5720			7922			10173		
# variables binaires	4294			4864			6285		
# contraintes	6156			10591			13366		

(section 3.2.4). La figure 5.4 illustre le séquençement temporel des sessions du profil de trafic 3 de l'instance (01)3-16.

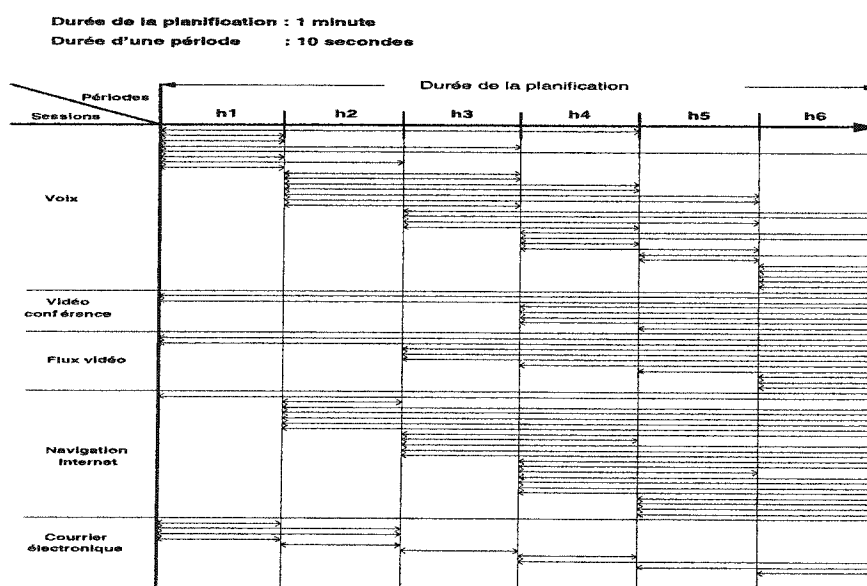


FIG. 5.4 – Séquençement temporel des sessions dans l'instance (01)3-16

### 5.3 Stratégies pour la résolution des problèmes

Dans cette section, nous comparons plusieurs combinaisons des stratégies de résolution de CPLEX-MIP pour trouver la plus adaptée à notre problème. Le choix des stratégies d'exploration, de séparation et de branchement est limité à celles proposées par CPLEX-MIP telles que décrites dans la section 5.1.3. De plus, il est possible d'affecter des priorités à certains groupes de variables dans les stratégies de séparation mais avec quelques restrictions : il n'est pas permis de donner des priorités en fonction des séparations précédentes. Ainsi, nous aurions voulu, par exemple, séparer suivant les variables  $\alpha_{irf}$  associées à la session  $s$  après que la procédure ait séparé suivant la variable  $y_s$ . Ceci est impossible avec le système de priorité de CPLEX-MIP où la seule possibilité est d'affecter des priorités globales aux variables sans pouvoir imposer des conditions pour les hiérarchiser au gré du processus de séparation. En nous accommodant de ces restrictions, nous proposons deux façons d'affecter les priorités :

#### – Priorité 1

Nous donnons des priorités uniquement aux variables  $y_s$  et  $x^{BS_i}$ . Les autres variables seront donc traitées comme si elles avaient la priorité la plus faible. Les variables  $y_s$  ont une priorité proportionnelle au chargement relatif, en termes de trafic, de la période qui correspond au début de leur requête. Ainsi, les variables  $y_s$  des sessions qui font une requête de connexion pendant la période la plus chargée ont la priorité la plus élevée. De même, les variables  $y_s$  des sessions qui font une requête pendant la période la moins chargée sont les moins prioritaires. Les variables  $y_s$  des sessions commençant à la même période ont la même priorité. En revanche, les variables  $x^{BS_i}$  ont toutes la même priorité. Cette priorité est, dans cette méthode, inférieure à la plus petite priorité des variables  $y_s$ . La procédure d'optimisation branchera donc d'abord sur les variables  $y_s$  dans l'ordre de leurs priorités, puis sur les variables  $x^{BS_i}$  si le choix se pose. Les autres variables seront traitées après.

#### – Priorité 2

Cette méthode est la même que la première excepté que, dans ce cas, la priorité des variables  $x^{BS_i}$  est plus élevée que la plus haute priorité des variables  $y_s$ . Ainsi, la procédure d'optimisation branchera d'abord sur les variables  $x^{BS_i}$  puis sur les

variables  $y_s$ .

L'affectation des priorités pour les variables  $y_s$  et  $x^{BS_i}$  se justifie par la structure du problème. En effet, les variables  $y_s$  et  $x^{BS_i}$  ont de forts impacts sur les autres variables. Quand une variable  $y_s$  est à zéro, les contraintes de CAC, couplées aux contraintes (4.28), (4.29), (4.24), (4.22), (4.12), (4.13), (4.14) et (4.20) forceront les variables  $\alpha_{irf}$ ,  $w_{irf}$  et  $z_{fr}$  à prendre la valeur entière zéro. Autant de possibilités en moins pour les prochains branchements du sous-arbre. De même, quand la variable  $y_s$  est à 1, pour chaque ensemble de variables binaires  $\alpha_{irf}$  et  $w_{irf}$ , associé à une session, une ou deux variables seront forcées à 1 par les contraintes (4.27), (4.28), (4.29) et (4.24). On privilégie le branchement sur la valeur 1 car les taux de rejet sont faibles et les solutions optimales auront une majorité de variables  $y_s$  à 1.

Quand la variable  $x^{BS_i}$  est à 1, elle force des variables  $\alpha_{irf}$  et  $w_{irf}$  à être égales à 1 par les contraintes (4.27), (4.28) et (4.34). Pour une session quelconque, ses variables  $\alpha_{irf}$  et  $w_{irf}$  seront choisies sur cette station de base  $i$  plutôt qu'une autre car l'objectif est de minimiser le nombre de stations de base. Cette méthode va permettre de réduire les possibilités de branchement rendues nombreuses avec ces variables binaires à triple indices. Pour les mêmes raisons que la priorité 1, on privilégie le branchement sur la valeur 1 pour les variables  $y_s$  et  $x^{BS_i}$ .

Le tableau 5.15 décrit les stratégies pour les cinq combinaisons les plus performantes parmi celles testées. La combinaison BR1 regroupe les stratégies par défaut de CPLEX-MIP. Les paramètres qui ne sont pas cités dans le tableau sont systématiquement pris avec leur valeur par défaut. Le critère de performance d'une combinaison BR est le temps CPU de la résolution ainsi que le nombre de nœuds explorés. Le tableau 5.16 donne, pour cinq jeux de données, les performances de chaque combinaison.

TAB. 5.15 – Combinaisons des stratégies de branchement

	BR1	BR2	BR3
CPX.PARAM.CLIQUES	normal	agressif	agressif
CPX.PARAM.COVERS	normal	agressif	agressif
CPX.PARAM.GUBCOVERS	normal	agressif	agressif
CPX.PARAM.NODESEL	Meilleures-Bornes	Meilleures-Bornes	Meilleure estimation
CPX.PARAM.VARSEL	automatique	automatique	automatique
CPX.PARAM.BRDIR	automatique	SUP	SUP
Priorité	non	non	non
CPX.PARAM.EPAGAP	$10^{-6}\%$	-	-
	BR4	BR5	
CPX.PARAM.CLIQUES	agressif	agressif	
CPX.PARAM.COVERS	agressif	agressif	
CPX.PARAM.GUBCOVERS	agressif	agressif	
CPX.PARAM.NODESEL	Meilleures-Bornes	Meilleures-Bornes	
CPX.PARAM.VARSEL	prédicatif	prédicatif	
CPX.PARAM.BRDIR	SUP	SUP	
Priorité	1	2	
CPX.PARAM.EPAGAP	-	-	

TAB. 5.16 – Performances des combinaisons de stratégies de branchement

	Instance (01)1-8	Instance (01)1-30	Instance (01)2-16	Instance (01)3-16	Instance (01)4-20
<b>BR1</b>	19.25 sec	3 heures	5 heures	2 jours	5 jours
	écart : $10^{-6}\%$	2 %	5 %	15.54 %	23.76 %
	313 nœuds	- nœuds	312 nœuds	- nœuds	- nœuds
<b>BR2</b>	7.66 sec	2 heures 20 min	24 min	2 jours	5 jours
	0.05 %	2 %	5 %	22.45 %	26.73 %
	86 nœuds	9825 nœuds	243 nœuds	- nœuds	- nœuds
<b>BR3</b>	23.31 sec	1 heure 15 min	15 min	40 heures	19 heures
	0.05 %	2 %	5 %	10 %	10 %
	275 nœuds	4044 nœuds	148 nœuds	7059 nœuds	4670
<b>BR4</b>	32.34 sec	25 heures	30 min	2 jours	4 jours
	0.05 %	7.25 %	5 %	29.74 %	31.65 %
	443 nœuds	- nœuds	184 nœuds	- nœuds	- nœuds
<b>BR5</b>	8.88 sec	1 heure $\frac{1}{2}$	35 min	23 heures	3 jours
	0.05 %	2 %	5 %	10 %	6.46 %
	89 nœuds	1891 nœuds	46 nœuds	9031 nœuds	7683 nœuds

On peut voir que les temps de résolution deviennent très grands quand le nombre de stations de base potentielles augmente (instance (01)1-8 et instance (01)1-30). En effet, les variables binaires, bien plus nombreuses pour l'instance (01)1-30, enrichissent l'arbre de branchement ce qui pénalise les temps de résolution. De plus, on remarque que le profil de trafic peut agir sur les temps de résolution, comme illustré avec les instances (01)2-16 et (01)3-16. La première instance contient 80 sessions de voix et prend moins de temps pour la résolution que l'instance (01)3-16 qui contient 80 sessions de diverses applications. La

différence entre ces deux profils repose plus sur la nature des contraintes que sur leur nombre ou sur le nombre de variables. Pour les applications VBR, les contraintes de CAC et les contraintes pour relier les débits entre la partie câblée et la partie radio sont des inégalités alors que pour les applications CBR ce sont des égalités. L'avantage des égalités est qu'elles réduisent l'espace des solutions permettant en général une résolution plus rapide.

D'après les résultats du tableau, on conclut que les stratégies BR3 et BR5 sont globalement les plus performantes pour la structure de nos problèmes. Pour d'autres jeux de données importants comme (02)5-25, la stratégie de branchement sur les nœuds de meilleure estimation est parfois inefficace en faisant grossir l'espace mémoire au fur et à mesure du parcours de l'arbre avec ses calculs d'estimation. La procédure était alors souvent arrêtée par CPLEX-MIP quand l'espace mémoire limite géré par le logiciel était dépassé.

Les stratégies BR3 et BR5 sont ainsi utilisées de facto pour les prochaines résolutions.

## 5.4 Validation du modèle

### 5.4.1 Validation de la contrainte qui relie les débits des flots d'une session au niveau de la transition entre le réseau de coeur et le réseau d'accès radio

La contrainte (4.22) qui relie le débit du réseau de coeur au débit du réseau d'accès radio est détaillée à la section 4.4.5. Cette contrainte force la session à avoir le même débit sur tous les liens câblés de sa route du nœud d'accès externe jusqu'à la ou les stations de base qui la prennent en charge. De plus, en cas de *soft* handover entre les stations de base  $i_1$  et  $i_2$ , on souhaite que :

$$x_{f_s^{wl}} = x_{f_s^{wl}}^{i_1} = x_{f_s^{wl}}^{i_2}.$$

Compte tenu du fait que la procédure d'optimisation minimise les débits  $x_{f_s^{wl}}^{i_1}$  et  $x_{f_s^{wl}}^{i_2}$ , elle affectera à ces deux variables le débit minimal imposé par la contrainte (4.20). Elles auront donc la même valeur. De plus, la procédure permettra d'obtenir une solution satisfaisant :

$$x_{f_s^{wl}} = \max_{(i \in L_s^{bs})} (x_{f_s^{wl}}^i).$$

D'après les résultats, certains débits n'étaient pas à leur valeur minimale. Parce que nous doutions de l'optimalité de la solution rendue, nous avons modifié la solution en injectant les débits attendus aux variables qui posaient problème pour se rendre compte finalement que



la fonction objectif donnait la même valeur que celle de la solution initiale. Rappelons que la fonction objectif (4.3) somme les capacités des liens câblés ainsi que le coût des stations de base sélectionnées. Or, la capacité d'un lien câblé obtenue dans une solution optimale est égale à la bande passante occupée par les sessions à la période la plus chargée pour le lien. Ainsi, si une session occupe le lien à une période moins chargée, l'objectif optimal ne sera pas changé même si son débit n'est pas au débit minimal. En revanche, toutes les sessions qui occupent le lien pendant sa période la plus chargée ont leurs débits affectés à la valeur minimale. Dans le sens montant, le RAB est supérieur au débit affecté sur les liens câblés (contrainte (4.13)), le débit choisi influence donc le chargement sur les liens radio. Cependant, même si le débit n'est pas minimal, il sera quand même choisi de façon à ne pas sélectionner une station de base supplémentaire, sinon l'objectif serait changé. Le modèle se comporte ainsi conformément à nos attentes pour les périodes critiques au dimensionnement.

En partant de la solution optimale rendue, il est possible de construire une autre solution optimale qui, en revanche, respecte la contrainte quadratique (4.21) pour toutes les sessions. Cette possibilité pourrait être à envisager si le modèle est étendu avec de nouvelles contraintes qui utiliseraient les valeurs des débits attribués aux flots.

#### 5.4.2 Influence du coût des stations de base sur le dimensionnement

La fonction objectif est la somme des capacités optimales des liens câblés du réseau et le nombre de stations de base. Les coûts des stations de base ainsi que les coûts des liens câblés dans la fonction objectif sont liés aux différentes tarifications pour l'achat, l'installation et la maintenance. Nous allons montrer dans cette section que suivant la charge de trafic, ces coûts ont un impact sur le dimensionnement.

En partant du principe que le coût des liens câblés est 1, le coût des stations de base devient le paramètre variable qui fixe le poids relatif des stations de base dans le dimensionnement. Pour comprendre l'influence de ce coût, nous proposons d'étudier l'instance (01)1-8 du tableau 5.13 en faisant varier le coût des stations de base. Certes, cette instance a un nombre de sessions réduit mais elle nous permet d'obtenir des solutions avec des écarts entre les bornes inférieure et supérieure de la solution optimale suffisamment faibles pour pouvoir tirer des conclusions sur le comportement du modèle. Le tableau 5.17 résume les résultats

obtenus pour cette instance après avoir fixé le coût des stations de base à 10, 500, 5000 puis 15000. Pour chaque coût, sont détaillés le nombre de stations de base sélectionnées,  $\sum BS_k$ , la somme des capacités optimales des liens câblés du réseau de cœur,  $\sum C_\ell$ , et la somme des capacités optimales des liens câblés du réseau d'accès radio,  $\sum C_{\ell_{BS_k}}$ . L'écart pour ces trois solutions est 0.05%, ce dernier, compte tenu des valeurs optimales de l'objectif et des différents coûts des stations de base, est suffisamment faible pour garantir que le nombre de stations de base sélectionnées est minimal.

TAB. 5.17 – *Influence du coût des stations de base*

Coût BS	Instance (01)1-8		
	$\sum BS_k$	$\sum C_\ell$	$\sum C_{\ell_{BS_k}}$
10	7	5583.2	1855.2
500	6	5583.2	1871.2
5000	6	5583.2	1871.2
15000	6	5583.2	1871.2

Les résultats du tableau 5.17 indiquent que le coût des stations de base influence les priorités du dimensionnement. En effet, quand le coût est égal à 10, le nombre de stations de base activées est plus élevé que pour les coûts de 500, 5000 et 15000. Le couplage se situe au niveau des liens câblés du réseau d'accès radio : le coût de la station de base supplémentaire est compensé par la réduction des capacités optimales  $\sum C_{\ell_{BS_k}}$ . En fait, cette station de base supplémentaire permet d'améliorer les capacités des liens câblés du réseau d'accès radio et ce, malgré le lien câblé ajouté entre la station de base supplémentaire et son BSC. En étudiant de plus près la solution optimale, on se rend compte que cette station de base a permis à la procédure de mieux répartir les sessions sur les stations de base et donc sur les liens du réseau d'accès radio pour en minimiser les capacités. En effet, les périodes de chargement maximal pour les liens câblés ne coïncident pas forcément. Ainsi, la procédure a intérêt à rattacher un usager mobile à une autre station de base si la session est active pendant la période de chargement maximal du lien rattaché à la première station de base et si, d'un autre côté, la session n'a pas d'impact sur la période la plus chargée du nouveau lien occupé. Sinon, l'opération est nulle car la capacité occupée par la session interviendra de la même façon sur le coût final.

En revanche, la part des liens câblés du réseau de cœur est la même pour les trois tests. Ceci est surprenant car un raisonnement similaire au précédent pourrait expliquer que les

capacités optimales des liens câblés du réseau de cœur soit plus faibles avec un coût de 10 pour les stations de base. En effet, une station de base supplémentaire donne plus possibilités à la procédure pour accepter ou refuser les sessions. Ainsi, elle peut mieux répartir les refus sur la période de planification pour minimiser l'ensemble des liens câblés du réseau. Pour cela, elle doit composer avec les périodes de chargement des liens et les routes des sessions. Dans notre cas, le nombre trop faible de périodes et de sessions, couplé avec des taux d'acceptation élevés, réduisent les possibilités d'améliorer les capacités : le détail des solutions optimales indiquent que les taux de refus par période et par application sont les mêmes pour les trois tests.

Plus le coût des stations de base est élevé, plus le couplage entre les liens câblés et les stations de base est faible, la priorité pour la procédure devenant le nombre de stations de base activées. Notons que passé un certain seuil, l'augmentation du coût n'apporte rien de nouveau pour le dimensionnement. Pour notre instance de trafic, le coût de 500 suffit à la procédure pour privilégier les stations de base. Les coûts de 5000 et 15000 ne permettent pas de réduire au delà de 6 le nombre de stations de base. Cependant, l'influence du coût des stations de base dépend de la charge de trafic. Si on veut conserver la priorité sur le nombre de stations de base, le coût de ces dernières doit nécessairement augmenter avec la charge de trafic. En effet, plus la charge de trafic est importante, plus les gains sur les capacités optimales des liens câblés seront importants car la procédure aura plus de choix pour répartir les sessions selon les périodes de chargement maximal des liens. Ainsi, la procédure sélectionnera une autre station de base pour réduire les coûts sur les liens câblés.

Pour un dimensionnement réaliste, les coûts des stations de base et des liens câblés sont directement liés à leur prix d'achat, à leur coût d'installation et à leur coût de maintenance. Suivant la charge de trafic, le couplage sera donc plus ou moins important, ce qui pourrait amener des dimensionnements de nature différente : quand le couplage est fort, les liens câblés seront privilégiés, ce qui entraînera une sous-utilisation des stations de base sélectionnées; avec un couplage faible, le nombre de stations de base sera mis en avant et leurs surfaces de couverture optimisées.

### 5.4.3 Impact du nombre de stations de base potentielles sur le dimensionnement

Plus le nombre de stations de base potentielles est élevé, plus les opportunités de trouver une station de base bien placée sont grandes. Ainsi, la procédure a plus de choix pour maximiser les rayons de couverture des stations de base et pour, éventuellement, en réduire le nombre. Cette tendance doit être combinée avec un couplage suffisamment faible pour que la procédure privilégie le nombre de stations de base activées. Ce cas s'est illustré avec le profil de trafic à 60 sessions pour laquelle la configuration initiale est de 30 stations de base potentielles (instance (01)1-30). Le tableau 5.18 donne le nombre de stations de base activées en fonction du nombre de stations de base potentielles. Les résultats sont obtenus avec un écart de 2 %. Les valeurs des objectifs obtenus sont tels que 2% de ces valeurs sont inférieurs au coût d'une station de base, ce qui garantit que nous avons le nombre minimal de stations de base activées.

TAB. 5.18 – *Impact du nombre de stations de base potentielles sur le dimensionnement pour le profil de trafic 1 (60 sessions)*

	60 sessions			
Nombre de stations de base potentielles	8	20	30	50
Nombre de stations de base sélectionnées (Coût BS = 500)	6	6	5	5

Les configurations à 30 et à 50 stations de base potentielles ont permis à la procédure de sélectionner une station de base en moins. De la même façon que pour le coût, la multiplicité des stations de base potentielles ne permet pas de réduire le nombre de stations de base activées indéfiniment, car, les limites de capacités radio imposent un nombre minimal de stations de base.

Nous proposons de détailler les solutions obtenues pour les deux instances dont la configuration initiale est de 20 et 30 stations de base potentielles.

Pour une même instance de trafic et une même topologie de réseau, quand on augmente le nombre de stations de base potentielles, les sous-ensembles  $L_s^{\text{BS}}$  et  $L_s^{\text{BS-SH}}$ , définis pour chaque usager mobile et qui influencent au final le dimensionnement optimal, changent (section 5.2.3). Pour que l'intérêt d'une configuration initiale riche en stations de base

potentielles soit conservé, il faut s'assurer que l'augmentation du nombre de stations de base potentielles enrichisse les choix de la procédure. Il faut donc que les sous-ensembles  $L_s^{\text{BS}}$  et  $L_s^{\text{BS-SH}}$  contiennent plus de stations de base. Pour cela, les distances  $d_{\text{BS-NON-SH}}$  et  $d_{\text{BS-SH}}$ , définies à la section 5.1.1.3, sont identiques d'une instance à l'autre quand on augmente le nombre de stations de base potentielles. Le tableau 5.19 récapitule les paramètres caractéristiques des deux instances et les résultats des solutions optimales. En outre, les nombres moyens de stations de base potentielles contenues dans les ensembles  $L_s^{\text{BS}}$  et  $L_s^{\text{BS-SH}}$ , ainsi que les distances  $d_{\text{BS-NON-SH}}$  et  $d_{\text{BS-SH}}$  sont précisés.

TAB. 5.19 – Configurations optimales pour le profil de trafic 1 (60 sessions)

	(01)1-20	(01)1-30
Nombre de stations de base potentielles	20	30
$L_s^{\text{BS}}$	11	14
$L_s^{\text{BS-SH}}$	9	11
$d_{\text{BS-NON-SH}}$	30 mètres	30 mètres
$d_{\text{BS-SH}}$	200 mètres	200 mètres
Objectif ( écart : 2% )	10438.4	9954.4
Nombre de stations de base sélectionnées	6	5
Liste des stations de base sélectionnées	5, 7, 9, 16, 17 et 19	11, 13, 15, 22 et 29

La figure 5.5 illustre à titre indicatif la disposition des stations de base sélectionnées pour les deux instances et fait apparaître les positions des usagers mobiles.

Le dimensionnement obtenu avec l'instance (01)1-30 supporte la même charge de trafic que l'instance (01)1-20 mais avec une station de base en moins. Indépendamment du coût final, les positions des stations de base de l'instance (01)1-30, détaillées sur la figure 5.5, optimisent donc les surfaces de couverture par rapport aux positions obtenues avec l'instance (01)1-20.

#### 5.4.4 Optimalité des positions des stations de base

L'optimalité des positions des stations de base est relative à l'ensemble discret des positions des stations de base potentielles. Nous venons de voir que le nombre de stations de base potentielles pouvait améliorer l'optimalité des positions pour les surfaces de couverture.

La discrétisation des stations de base potentielles permet un quadrillage qui évite les positions non réalisables sans contrainte supplémentaire. En outre, il est possible de pri-

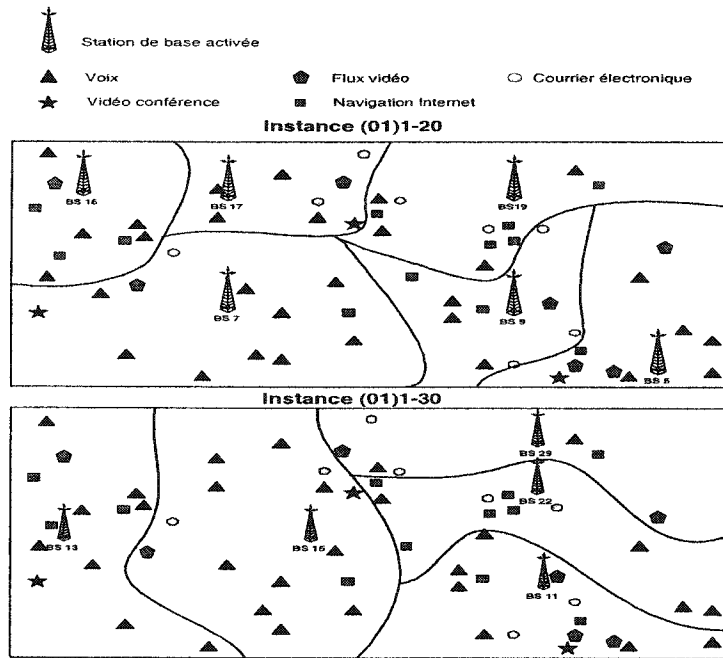


FIG. 5.5 – Positions des stations de base activées dans les instances (01)1-20 et (01)1-30

vilégier certaines positions avec des coûts d'installation plus faibles. Un autre intérêt de la discrétisation est de limiter le nombre de variables et de conserver des contraintes linéaires. En effet, avec des variables continues pour les positions des stations de base, les contraintes de capacité radio seraient quadratiques : ces variables continueraient d'intervenir dans les formules de capacité radio au niveau des coefficients d'atténuation  $a_{m,0}$  et  $a_{m,i}$ . En appliquant la formule (3.10), la distance  $d$  deviendrait une fonction de ces variables continues. La surface à dimensionner peut être découpée en plusieurs zones de trafic homogène dont les densités de trafic seraient variables. A l'image du modèle<sup>[17]</sup>, certaines zones de trafic seraient privilégiées en imposant des coûts d'installation plus faibles pour les stations de base réparties sur ces zones. De plus, plusieurs groupes de stations de base potentielles seraient définis dans chaque zone selon qu'elles sont installées sur les murs d'un immeuble, en haut d'une tour ou autres. Ces groupes de stations de base auraient des coûts différents. Cependant, les formules de capacité radio se basent sur l'hypothèse d'un trafic homogène (sections 3.1.5.3 et 3.1.5.3), et ne peuvent donc s'appliquer que pour une de ces zones de trafic définies précédemment. Ainsi, la surface de couverture testée dans nos instances se limite à une zone de trafic homogène. C'est aussi pour cette raison que les stations de base

potentielles sont initialement réparties de façon uniforme (section 3.5.1). De plus, nous supposons que nous dimensionnons uniquement un groupe de stations de base en attribuant les mêmes coûts à toutes les stations de base potentielles.

La question de l'optimalité des positions sélectionnées s'est posée avec les instances (01)1-8 et (01)1-20 pour lesquelles nous avons obtenu plusieurs solutions au même coût optimal. Pour cela, après les premières solutions, nous avons relancé une deuxième procédure après avoir forcé le modèle à sélectionner une station de base particulière ( $BS_k$  forcée à 1). Le tableau 5.20 indique les stations de base sélectionnées pour ces différents tests.

TAB. 5.20 – *Combinaisons des stations de base activées à partir d'une même configuration initiale*

	(01)1-8		(01)1-20	
Nombre de stations de base potentielles	20		30	
Station de base forcée	-	#7	-	#4
Objectif	1054.4 (écart : 0.05%)	1054.4 ( 0.05% )	10438.4 ( 2% )	10438.4 ( 2% )
Nombre de stations de base sélectionnées	6	6	6	6
Liste des stations de base sélectionnées	2, 3, 5, 6, 7 et 8	2, 4, 5, 6, 7 et 8	5, 7, 9, 16, 17 et 19	2, 4, 11, 12, 13 et 19

On peut voir, pour ces deux cas de figure, qu'il est en effet possible de trouver un autre combinaison de stations de base activées avec le même coût optimal.

Il pourrait être intéressant d'enrichir le critère d'optimalité des positions à l'intérieur d'une zone de trafic homogène pour que la procédure puisse choisir une combinaison plutôt qu'une autre. Ce critère pourrait être la répartition du taux d'occupation sur les stations de base. En effet, supposons qu'une station de base soit en limite de charge maximale avec le dimensionnement rendu. Une fois ce dimensionnement implanté, cette station de base aura plus de chance d'alourdir les taux de rejet lorsque le trafic réel s'écartera légèrement du trafic escompté. Ainsi, une répartition plus homogène sur l'ensemble des stations de base permettrait de généraliser plus efficacement le dimensionnement aux variations du trafic réel par rapport à l'instance de trafic testée. Ce critère n'a de sens que si le trafic est supposé homogène sur l'aire géographique. La figure 5.6 décrit, pour les deux instances du tableau 5.20, les taux d'occupation des stations de base sélectionnées dans chaque combinaison. Les taux d'occupation sont donnés pour les trois périodes les plus chargées (sur les 6 composant

la durée de planification). Pour chaque diagramme, on indique l'écart observé correspondant à la différence entre les taux d'occupation de la station de base la plus chargée et la moins chargée. En suivant le raisonnement précédemment, le critère d'optimalité des positions porterait sur cet écart.

D'après les écarts observés, la meilleure solution pour les deux configurations à 8 et 20 stations de base potentielles correspond aux instances avec une station de base forcée. Le choix n'est pas vraiment tranché pour ces deux exemples mais il existe possiblement d'autres combinaisons encore meilleures.

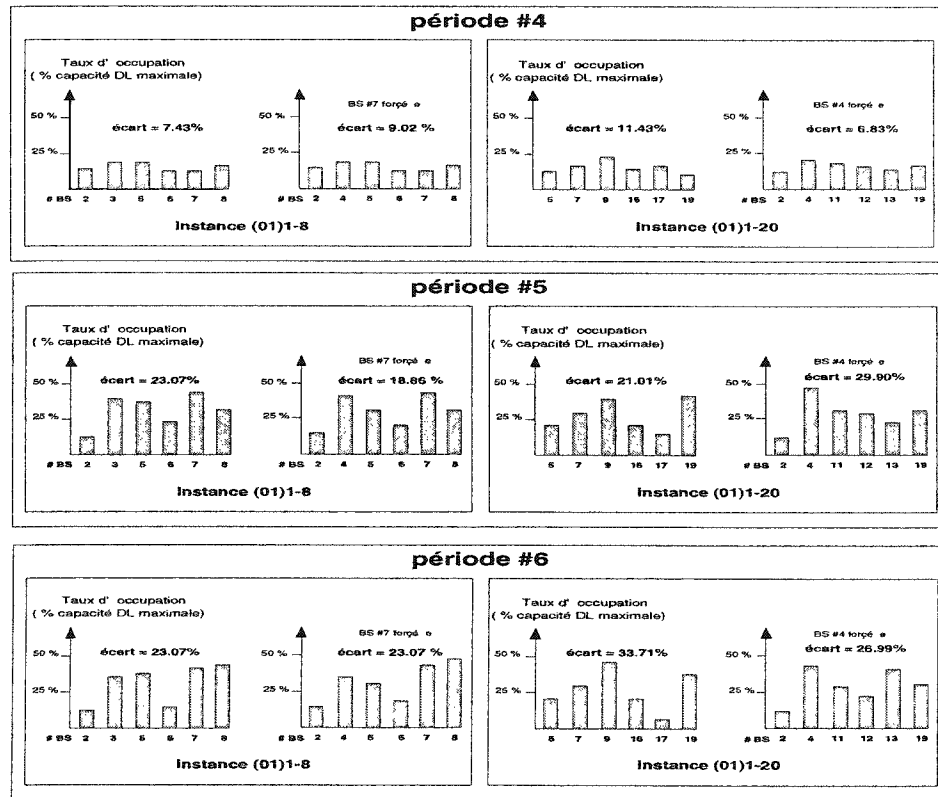


FIG. 5.6 – Taux d'occupation des stations de base sur les périodes les plus chargées

Le nombre de stations de base potentielles peut avoir un effet sur ce critère d'optimalité car il multiplie les combinaisons des stations de base sélectionnées qui amènent au même coût optimal. En effet, plus les stations de base potentielles sont rapprochées, moins on a de chance de modifier les taux de refus et les débits sélectionnés en passant



d'une station de base à une autre voisine. Cependant, les capacités radio occupées changent systématiquement, et les taux d'occupation avec. Ainsi, la multiplicité des stations de base potentielles permettrait d'améliorer ce critère en offrant plus de possibilités. Mais ce, sans véritable garantie non plus.

Avec une configuration où plusieurs groupes de stations de base sont représentés (coûts différents dans la fonction objectif), les possibilités de multiples combinaisons s'appliqueront toujours à l'intérieur des groupes mais deviendront probablement plus limitées pour la configuration globale.

#### 5.4.5 Le *soft handover*

Le *soft handover* ajoute un degré de liberté au couplage liens câblés-stations de base évoqué à la section 5.4.2. En effet, accepter une session en *soft handover* permet à la procédure de réduire la capacité radio requise et d'accepter la session sans ajouter une nouvelle station de base. D'un autre côté, la session sera plus pénalisante pour les liens câblés du réseau d'accès radio car, au lieu d'un lien, elle va occuper deux liens : un entre chaque station de base et son BSC.

En suivant ce raisonnement, on en déduit que le pourcentage des sessions en *soft handover* devrait être nul quand le coût des stations de base est très faible. La procédure choisira dans ce cas de rajouter une station de base plutôt que d'augmenter la capacité des liens câblés du réseau d'accès radio. Le tableau 5.21 récapitule le pourcentage des sessions acceptées en *soft handover* pour plusieurs instances.

TAB. 5.21 – *Taux d'utilisateurs en soft handover*

	(01)1-8		(01)1-20	(01)3-16	(01)4-20
Nombre de sessions	60		60	80	150
Coût BS	10	500	500	500	500
Ecart	0.05%	0.05 %	2%	10 %	6.46 %
Objectif	7508.4	1054.4	10438.4	22368.8	24063.2
Taux d'utilisateurs en <i>soft handover</i>	3.846 %	5.769 %	5.769 %	14.705 %	9.836 %

Si on s'arrête à ces résultats, le comportement du *soft handover* ne répond pas vraiment à nos attentes. Malgré le coût des stations de base à 10, pratiquement 4% des sessions sont en *soft handover*. En revanche, les pourcentages sont quand même plus importants avec des profils de trafic chargés. Pour approfondir le comportement du modèle, nous avons relancé

certaines instances en ajoutant une contrainte au problème. L'objectif de cette contrainte est de forcer le pourcentage des sessions en *soft* handover à être supérieur ou égal à une valeur choisie  $P_{SH}$ . Pour exploiter les résultats, il faut être capable de résoudre l'instance avec un écart suffisamment petit. Nous choisissons donc les instances (01)1-8 et (01)1-20. La forme de la contrainte pour imposer un taux minimal est la suivante :

$$\sum_{s \in S} \sum_{r \in T_s^{ul} \text{ ou } r \in T_s^{dl}} z_{fr} \geq P_{SH} \sum_{s \in S} y_s. \quad (5.1)$$

La somme des variables  $z_{fr}$  se fait sur l'ensemble des RABs soit pour le sens montant soit pour le sens descendant. En effet, si la session est en *soft* handover, on a :

$$\sum_{r \in T_s^{ul}} z_{fr} = \sum_{r \in T_s^{dl}} z_{fr} = 1.$$

La fraction dans la contrainte (5.1) représente ainsi le rapport du nombre de sessions en *soft* handover sur le nombre de sessions acceptées.

Nous avons relancé l'instance (01)1-8 avec le coût de 10 pour les stations de base et en forçant le pourcentage de *soft* handover à être plus petit que 1% (inégalité inversée dans la contrainte (5.1)). Le pourcentage effectivement obtenu est 0% et le coût optimal reste identique. De la même façon, nous avons forcé le pourcentage à être supérieur à 10%, 20% puis 50 % pour l'instance (01)1-20. C'est en forçant le taux à 50% que la solution optimale est moins bonne que les précédentes. De même que pour les positions des stations de base, il est donc possible de trouver plusieurs combinaisons pour le *soft* handover. Si le chargement des deux stations de base le permet, la procédure peut choisir d'accepter l'utilisateur mobile avec ou sans *soft* handover, même si cette dernière est dans une zone de recouvrement commune aux deux stations de base. Ce choix peut être sans conséquence sur le nombre final de stations de base sélectionnées. De plus, le coût supplémentaire sur les liens câblés peut aussi être sans effet sur la valeur de la fonction objectif si les sessions en *soft* handover sont réparties en dehors des périodes de chargement maximal des liens concernés.

Dans notre modèle, l'intérêt de la modélisation du *soft* handover est de permettre à la procédure d'optimiser le dimensionnement dans le cas particulier où les taux d'occupation

des stations de base seraient très élevés et que les usagers mobiles en *soft* handover pourraient éviter une station de base supplémentaire. En revanche, le modèle ne permet pas une disposition des sessions en *soft* handover nécessairement réaliste car il n'intègre pas vraiment de critère spatial. La seule contrainte que l'on impose est qu'un usager mobile ne peut pas être pris en *soft* handover par deux stations de base dont l'une est trop proche de l'utilisateur (distance  $d_{\text{BS-NON-SH}}$ ). Cependant, un usager peut se retrouver en *soft* handover entre deux stations de base sans être sur leur zone commune de recouvrement, et inversement un usager placé sur cette zone de transition n'est pas forcément pris en *soft* handover. La figure 5.7 illustre ces possibilités avec la disposition des usagers mobiles en *soft* handover obtenue pour l'instance (01)3-16. Dans un cas idéal sans interférence, le *soft* handover de l'utilisateur mobile marqué d'une étoile sur la figure 5.7 pourrait se produire si les stations de base plus proches étaient en limite de charge maximale, le seul objectif étant d'accepter la session malgré les chemins détournés. Cependant, ce type de *soft* handover est très néfaste pour les stations de base voisines car l'utilisateur mobile émet à une puissance élevée pour assurer les liaisons avec ses deux stations de base éloignées. Le niveau d'interférence inter-cellulaires pour les stations de base les plus proches augmente alors fortement, ce que le modèle ne prend pas en compte avec les formules de capacité radio car les interférences inter-cellulaires sont moyennées (section 3.1.5). Ce cas ne se produit probablement pas ou très peu dans la réalité, le *soft* handover s'appliquant principalement, dans un cas idéal de propagation, aux usagers qui se trouvent sur les zones de recouvrement des cellules pour limiter les interférences intra-cellulaire dans les deux cellules du *soft* handover et les interférences inter-cellulaires dans les cellules voisines.

## 5.5 Résultats pour le dimensionnement des réseaux 3G

### 5.5.1 Taux d'occupation des stations de base

Dans cette section, nous nous attardons sur l'évolution des taux d'occupation des stations de base pour trois profils de trafic :

- un profil de trafic P1 avec uniquement de la voix, correspondant à un profil de trafic pour les réseaux 2G,
- un profil de trafic P2 avec toutes les applications mais où la voix reste prédominante,

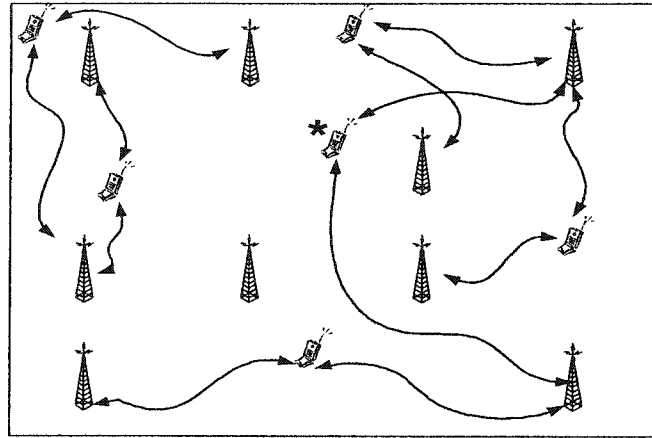


FIG. 5.7 – Répartition des usagers en soft handover

- un profil de trafic P3 avec toutes les applications mais où la navigation Internet, très exigeante en débit dans le sens descendant, est largement représentée.

Les instances testées se basent sur l'instance (01)3-16 : le nombre de sessions, la durée de planification, la topologie du réseau, le nombre de stations de base potentielles et les RABs proposés par application sont identiques. Seule change la distribution des sessions selon les applications. Le tableau 5.22 décrit la distribution des sessions pour les trois profils de trafic P1, P2 et P3, ce dernier correspondant au profil de trafic de l'instance (01)3-16.

TAB. 5.22 – Répartition des sessions selon les applications pour les profils de trafic P1, P2 et P3

	P1	P2	P3
Voix	80	35	30
Vidéo conférence	0	5	7
Flux vidéo	0	10	10
Navigation Internet	0	10	23
Courrier électronique	0	20	10
<b>Total</b>	80	80	80

La figure 5.8 décrit, pour les trois périodes les plus chargées, les taux d'occupation des stations de base pour le sens montant (UL) et descendant (DL). Les taux d'occupation sur une période sont évalués en appliquant la formule suivante :

$$\text{taux d'occupation (\%)} = 100 \times \frac{\text{Capacité radio occupée}}{\text{Capacité radio maximale}}$$

Les capacités maximales des stations de base ainsi que les capacités radio occupées par les sessions sont décrites aux sections 3.1.5.3 et 3.1.5.4, pour le sens montant et descendant

respectivement.

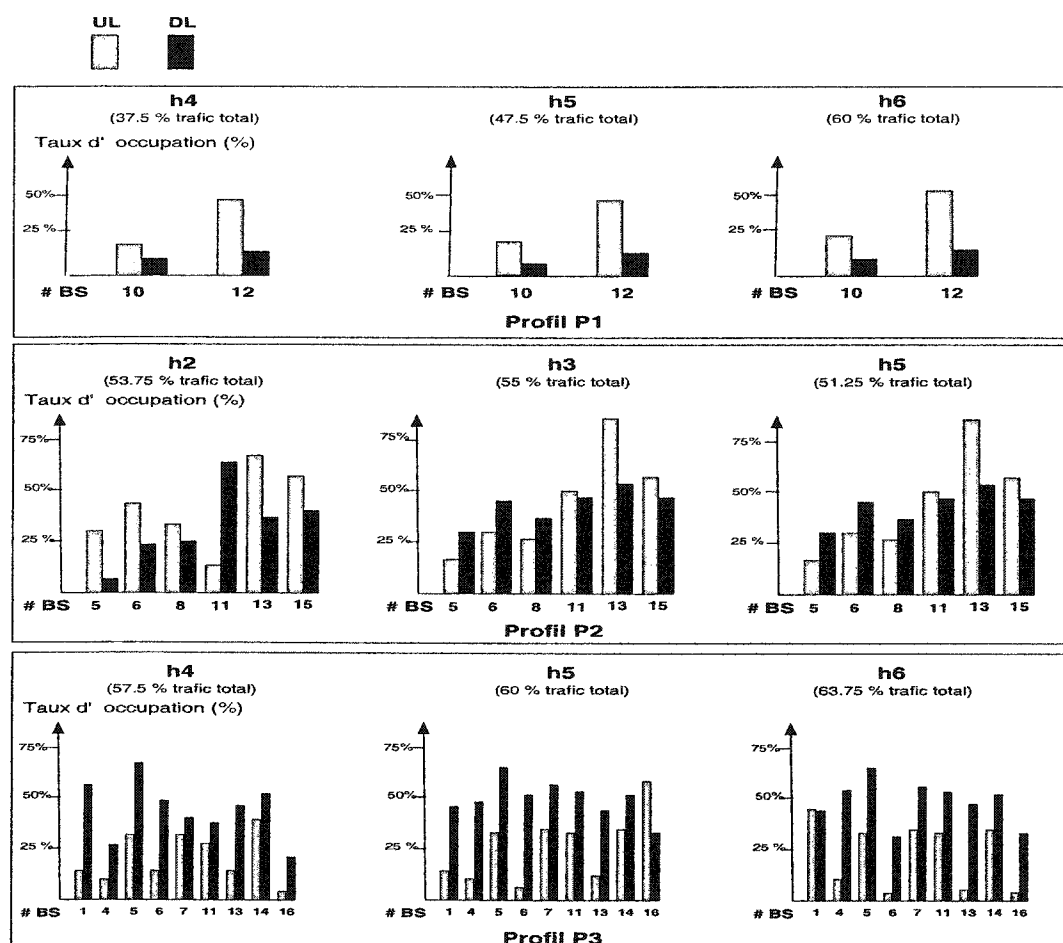


FIG. 5.8 – Evolution des taux d'occupation des stations de base suivant le profil de trafic

On remarque, d'après la figure 5.8, que les taux d'occupation sont plus importants dans le sens montant pour le profil P1. Puis, les taux d'occupation dans le sens descendant s'alignent avec ceux du sens montant pour le profil P2 et finalement, ils deviennent bien plus importants avec le profil P3. Cette observation illustre le revirement des réseaux 3G concernant les capacités limites dans les deux directions. En effet, pour les réseaux 2G, dont le trafic se résume essentiellement à la voix, la direction critique en termes de capacité radio est le sens montant. Ceci s'explique par les interférences intra-cellulaires qui, dans le sens montant, sont bien plus importantes que dans le sens descendant car l'orthogonalité des

codes, qui permet d'éliminer les interférences entre les usagers rattachés à une même cellule, est fortement dégradée dans le sens montant (section 3.1.5). Ainsi, avec des applications dont les flots montant et descendant sont symétriques, la direction montante sera plus vite saturée en raison des interférences intra-cellulaires. En revanche, avec des applications asymétriques qui demandent des débits élevés dans le sens descendant (profils de type P3), la direction critique devient le sens descendant, les capacités radio pour ces flots devenant bien plus pénalisantes que les interférences intra-cellulaire du sens montant. Ainsi, la capacité maximale des stations de base sera plus vite atteinte dans le sens descendant. Ces observations nous permettent de valider partiellement le comportement des formules de capacité radio utilisées dans notre modèle.

## 5.5.2 La qualité de service et les niveaux de service par application

### 5.5.2.1 La qualité de service

Dans notre modèle, la qualité de service intervient sous plusieurs formes :

- garantie sur le délai de bout en bout d'un paquet dans le sens descendant (contrainte (4.20)),
- garanties sur les débits minimaux pour l'ensemble des flots associés à une session (contraintes de CAC (4.6), (4.8), (4.7) et (4.9)),
- garanties statistiques sur le niveau de qualité de service avec l'attribution des RABs (contrainte (4.25)).

Un exemple de calcul des débits minimaux nécessaires pour garantir les exigences de délai des applications VBR est donné dans l'annexe III.

Nous proposons, dans cette section, d'observer les taux de répartition des sessions selon les RABs. Les coefficients  $Q_{r,a,p}$ , dans la contrainte (4.25), forcent la procédure à respecter des pourcentages de répartition selon les RABs. Prenons l'exemple de la navigation Internet High dans le sens descendant et rappelons les valeurs de RAB proposés avec leur coefficients  $Q_{r,a,p}$  associés.

D'après les explications sur la contrainte (4.25) données à la section 4.4.6, on s'attend à ce que, pour la navigation Internet, au moins 60 % des usagers Gold soient acceptés au RAB r1 du tableau 5.23. Ensuite, on souhaite qu'il y ait environ 25 % des usagers Gold acceptés

TAB. 5.23 – *Navigation Internet High*

Navigation Internet High				
	RAB $r$	$r_t$ (kbps); $r_{FER}$	$Q_{r,a,p}$	
			Gold	Silver
$T_s^{DL}$	r1	[307.2; 5%]	60%	10%
	r2	[307.2; 10%]	25%	5%
	r3	[153.6; 5%]	10%	60%
	r4	[153.6; 10%]	5%	25%

au RAB r2, 10 % au RAB r3 et 5 % au RAB r4. Pour ces trois derniers RABs, la contrainte (4.25) tolère des taux d'acceptation légèrement inférieurs, on privilégie cependant les RABs associés aux coefficients  $Q_{r,a,p}$  les plus élevés. Ainsi, il y aura toujours plus de sessions au RAB r2 qu'au RAB r3 ou r4 pour les usagers Gold. Les taux de répartition pour les usagers Silver privilégient les RABs impliquant une qualité de service moins bonne (débit plus faible et/ou FER plus élevé). Statistiquement, un usager Silver recevra donc un niveau de qualité de service inférieur à celui de l'utilisateur Gold.

Dans le cadre de notre étude, nous sommes capables d'obtenir des résultats avec des écarts raisonnables (10%) pour des instances où les ensembles  $T_s^{UL}$  et  $T_s^{DL}$ , définis pour chaque type d'application, sont réduits à deux RABs. Comme expliqué à la section 5.2.2, les coefficients  $Q_{r,a,p}$  pour les applications de navigation Internet et de courrier électronique sont, dans ce cas, changés en 90% et 10%.

Le tableau 5.24 regroupe les taux de répartition des sessions effectivement observés à partir de la solution optimale obtenue pour le profil  $P2$  décrit à la section précédente.

Parce que la procédure minimise les débits et les ressources radio, elle va chercher à exploiter au mieux les RABs les moins exigeants. Heureusement, les coefficients  $Q_{r,a,p}$  empêchent que toutes les sessions soient acceptées au RAB impliquant la qualité de service la moins bonne. Cependant, le tableau 5.24 indique que, pour la plupart des applications et des priorités, les sessions sont toutes acceptées au meilleur RAB ( $Q_{r,a,p}$  atteint=100%). Ceci s'explique par le nombre peu élevé de sessions d'un type donné et d'une priorité donnée. Le tableau 5.25 précise le nombre de sessions, par application et par priorité, acceptées dans la solution optimale.

TAB. 5.24 – Répartition des sessions selon les niveaux de qualité de service  $Q_{r,a,p}$  pour le profil P2 (80 sessions)

Qualité de Service					
		Voix Gold		Voix Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
UL -r1 :	9.6 kbps , 1% ]	90%	90.9091%	60%	64.7059 %
UL -r2 :	9.6 kbps , 2% ]	10%	9.09091%	40%	35.2941 %
DL -r1 :	9.6 kbps , 1% ]	90%	90.9091%	60%	64.7059 %
DL -r2 :	9.6 kbps , 2% ]	10%	9.09091%	40%	35.2941 %
		Vidéo conférence LOW Gold		Vidéo conférence Low Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
UL -r1 :	76.8 kbps , 0.5% ]	70%	100%	30%	0 %
UL -r2 :	76.8 kbps , 1% ]	30%	0%	70%	100 %
DL -r1 :	76.8 kbps , 0.5% ]	70%	100%	30%	0 %
DL -r2 :	76.8 kbps , 1% ]	30%	0%	70%	100 %
		Flux vidéo LOW Gold		Flux vidéo LOW Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
DL -r1 :	38.4 kbps , 0.5% ]	70%	100%	30%	22.2222 %
DL -r2 :	38.4 kbps , 1% ]	30%	0%	70%	77.7778 %
		Navigation Internet HIGH Gold		Navigation Internet HIGH Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
UL -r1 :	19.2 kbps , 5% ]	90%	100%	10%	0 %
UL -r2 :	19.2 kbps , 10% ]	10%	0%	90%	100 %
DL -r1 :	307.2 kbps , 5% ]	90%	100%	10%	0 %
DL -r2 :	307.2 kbps , 10% ]	10%	0%	90%	100 %
		Courrier électronique UL Gold		Courrier électronique UL Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
UL -r1 :	153.6 kbps , 5% ]	90%	100%	10%	0 %
UL -r2 :	153.6 kbps , 10% ]	10%	0%	90%	100 %
		Courrier électronique DL Gold		Courrier électronique DL Silver	
RAB [ $r_t$ (kbps), $r_{FER}$ (%) ]		$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint	$Q_{r,a,p}$ cible	$Q_{r,a,p}$ atteint
DL -r1 :	153.6 kbps , 5% ]	90%	100%	10%	0 %
DL -r2 :	153.6 kbps , 10% ]	10%	0%	90%	100 %

TAB. 5.25 – Répartition des sessions acceptées pour l'instance associée au profil P2

	Gold	Silver
Voix	11	17
Vidéo conférence LOW	3	2
Flux vidéo LOW	1	9
Navigation Internet HIGH	4	5
Courrier électronique UL	5	3
Courrier électronique DL	5	4

90% des 5 sessions de courrier électronique UL-Gold, par exemple, donne 4.5. La procédure est obligée de prendre les 5 sessions à ce RAB pour respecter la contrainte (4.25) associée au coefficient  $Q_{r,a,p}$  de 90% (r1), ce qui est possible car la contrainte de QoS associée au



second RAB r2 tolère un taux inférieur à 10 %. En revanche, pour les applications de Voix et de Flux vidéo Low Silver, le nombre de sessions est suffisamment élevé pour obtenir des taux de répartition satisfaisants.

### 5.5.2.2 Les niveaux de service

Le modèle intègre le concept des niveaux de service en forçant la procédure à accepter un quota minimum  $(1 - Br_{a,p})$  de sessions par application et par priorité (contrainte (4.10)). Ces quotas sont globaux à toute la durée de planification et la procédure répartit les refus sur toutes les périodes avec comme unique objectif la minimisation des ressources. Les quotas peuvent donc être violés sur une période tant que le niveau de service global est respecté. Les résultats sur les niveaux de service ne sont représentatifs que pour un nombre de sessions suffisamment important (environ 150 sessions). Nous proposons donc d'étudier l'instance (01)4-20 en tolérant un écart de 20% et pour laquelle les ensembles  $T_s^{UL}$  et  $T_s^{DL}$  sont réduits à un RAB. Le tableau 5.26 résume les niveaux de service globaux effectivement atteints.

TAB. 5.26 – Taux d'acceptation atteints avec l'instance (01)4-20 (150 sessions)

	Gold		Silver	
	GoS cible ( $1 - Br_{a,p}$ )	GoS atteint	GoS cible ( $1 - Br_{a,p}$ )	GoS atteint
<b>Voix</b>	90 %	93.75 %	70 %	70.4545 %
<b>Vidéo conférence LOW</b>	90 %	100 %	70 %	72.7273 %
<b>Flux vidéo LOW</b>	90 %	90.9091 %	70 %	77.7778 %
<b>Navigation Internet HIGH</b>	90 %	94.1176 %	70 %	76.9231 %
<b>Courrier électronique</b>	90 %	91.6667 %	70 %	76.9231%

Comme l'objectif est de minimiser les ressources, la procédure cherche à refuser le plus de sessions possibles. Ainsi, plus il y a de sessions, plus on tend vers les quotas cibles. Le principe de notre contrôle d'admission est anticipatif et utilise le profil de trafic sur toute la durée de planification pour décider de refuser ou accepter une requête. Les niveaux de service globaux lui permettent de répartir au mieux les refus et le modèle lui donne toute marge de manœuvre pour refuser une session même si les ressources sont disponibles au moment de la requête. La répartition des refus est ainsi optimale car elle se fait en fonction de la charge totale du trafic, de la durée des sessions et des liens occupés sans souci de causalité. La figure 5.9 illustre à titre indicatif le comportement de ce CAC anticipatif en détaillant les variations du niveau de service de la voix atteint sur les six périodes de la

planification. Le niveau de service pour une période donnée est le taux de refus toléré des sessions qui font leur requête à cette période. La voix a été choisie en particulier pour cette figure car les sessions de voix sont suffisamment nombreuses, dans cette instance, pour qu'à chaque période il y ait de nouvelles requêtes.

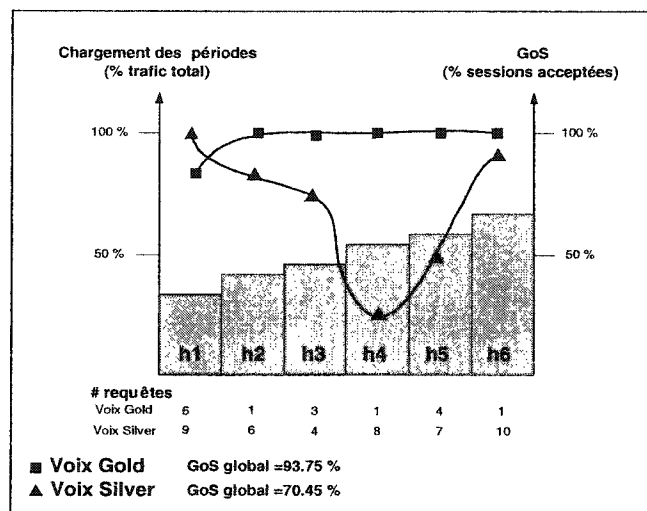


FIG. 5.9 – Variations sur la durée de planification du niveau de service atteint pour les applications de Voix

L'aspect anticipatif du CAC est illustré en particulier avec la courbe des Voix Silver : pour les deux dernières périodes où le chargement est le plus important, le taux d'acceptation augmente pour s'approcher du 90 %. Il y a deux raisons à ce comportement : compte tenu de la durée de planification (1 minute), la plupart des sessions qui font leur requête au début ont une durée de six périodes. Ainsi refuser une de ces sessions permet de libérer plus de ressources sur plusieurs périodes pour d'autres sessions qui devront être acceptées à cause du taux d'acceptation minimal. La deuxième raison est l'approche globale de la procédure qui tient compte de plusieurs facteurs (durée, route, qualité de service etc ...) pour accepter ou refuser une session sans souci de causalité. Pour un CAC causal, la courbe aurait eu une progression décroissante continue : les taux de rejet devenant plus importants quand le chargement global du trafic augmente. Le choix d'un tel CAC se base uniquement sur les ressources disponibles au moment de la requête : quand le chargement global du trafic augmente, les ressources sont plus rares et deviennent insuffisantes pour accepter toutes les requêtes.

### 5.5.3 Dimensionnement des réseaux 3G

Les réseaux 3G vont certainement se développer, dans un premier temps, dans des îlots de couverture (milieu urbain, centres d'affaires, etc.) et se généraliser par un déploiement progressif, permettant des investissements incrémentaux. Pour les premières phases du déploiement, la 3G s'appuiera donc sur les réseaux existants (2G) pour la couverture globale. L'objectif est d'obtenir une couverture maximale telle qu'en tous lieux, les services multimédia soient accessibles à haut débit dans les îlots 3G ou en mode dégradé lorsque la 2G prend le relais. Cela implique une interopérabilité maximale entre les deux systèmes, autant pour les réseaux que pour les terminaux mobiles, pour permettre aux usagers de passer d'îlot en îlot tout en respectant une certaine continuité de service.

La première étape au déploiement de la 3G sera donc la mise en place de cellules pour couvrir les zones confinées des centres villes où la demande en trafic sera importante. Nous proposons dans cette section, d'observer l'évolution du dimensionnement nécessaire pour supporter une demande de trafic croissante dans un tel îlot. L'espace étudié est une surface de  $500 \times 400$  mètres. Les profils de trafic pour les instances de 50 et 100 sessions suivent le scénario A décrit par le tableau 5.1. Ce scénario prédit l'utilisation des applications par les usagers quand la 3G commencera à s'implanter. Avec des jeux de données plus importants de 150 et 250 sessions, nous utilisons plutôt le scénario B qui s'appliquerait une fois que la 3G sera largement déployée (d'après [30]).

Pour résoudre ces profils de trafic chargés, nous réduisons les ensembles  $T_s^{UL}$  et  $T_s^{DL}$  à un seul RAB. La topologie du réseau est réseau02. Le cas à 150 sessions est l'instance (02)5-25 du tableau 5.14. Les autres jeux de données se basent sur la même topologie de réseau et les mêmes ensembles de RAB. Seuls changent le nombre de stations de base potentielles et les profils de trafic. Le tableau 5.27 donne, pour chaque instance, la valeur de ces différents paramètres variables.

TAB. 5.27 – *Instances pour le dimensionnement des réseaux 3G*

	50 sessions	100 sessions	150 sessions	250 sessions
Scénario	A	A	B	B
Nombre de stations de base potentielles	16	20	25	30
Voix	20	40	89	149
Vidéo conférence Low + High	2	5	7	12
Flux vidéo Low + High	6	10	14	22
Navigation Internet Low + High	12	25	15	25
Courrier électronique	10	20	25	42

Le tableau 5.28 indique pour chacune de ces instances le nombre de stations de base à déployer ainsi que le coût lié aux capacités des liens câblés. Ce coût représente la capacité totale en kbps sur l'ensemble du réseau.

TAB. 5.28 – *Dimensionnement des réseaux 3G en fonction de la charge de trafic*

	Scénario A		Scénario B	
	50 sessions	100 sessions	150 sessions	250 sessions
Ecart de la solution	20 %	27 %	31 %	33 %
Nombre de stations de base activées	9	11	24	30
Coût des liens câblés	11504.0	17745.6	21721.6	35347.2
Rayon moyen des cellules	84 mètres	76 mètres	67 mètres	56 mètres

Il apparaît clairement que les infrastructures des réseaux 3G seront nombreuses et coûteuses. Le tableau précise pour chaque instance le rayon moyen de couverture des stations de base. Ce rayon moyen est calculé en divisant l'aire géographique par le nombre de stations de base sélectionnées. Les rayons de couverture obtenus en calculant, pour chaque station de base, les distances d'éloignement des usagers mobiles rattachés sont plus élevés. Ceci s'explique par le fait que la procédure couple les liens câblés et les stations de base pour minimiser l'objectif. Ainsi, pour minimiser les liens câblés du réseau d'accès radio, la procédure peut rattacher un usager mobile à une station de base plus éloignée qu'une autre, aussi activée et qui dans la réalité aurait pris en charge l'utilisateur. Ainsi, ces rayons sont surestimés. C'est pourquoi nous préférons indiquer les rayons moyens en divisant la surface de couverture pour donner une idée plus réaliste des surface de couverture.

L'évolution des rayons moyens montre la nécessité de déployer des pico-cellules dont la portée est de l'ordre de la dizaine de mètres. En effet, parce que les réseaux 3G offrent des débits à la fois variables et importants, ils nécessitent des cellules de taille nettement plus petite que les macro-cellules actuelles, le facteur limitant étant les capacités radio dans le sens descendant pour applications exigeantes en débit (section 5.5.1). Cela conduira à un

réseau au coût plus élevé, onéreux en infrastructures.

#### 5.5.4 Conclusions

L'objectif de ce chapitre a été, d'une part, de présenter les différentes étapes depuis la formulation mathématique du modèle jusqu'à l'obtention des résultats, en détaillant la structure du programme, en donnant les valeurs des paramètres qui interviennent dans le modèle et en précisant les méthodes de résolution, et, d'autre part, de valider le modèle en étudiant son comportement par rapport à plusieurs instances. De plus, une étude de dimensionnement sur plusieurs chargements de trafic est menée sommairement dans la dernière partie.

#### Méthodes de résolution

Le choix de linéariser les contraintes a coûté en nombre de variables et en nombre de contraintes, mais cela nous a permis d'utiliser CPLEX-MIP pour résoudre des instances allant jusqu'à 200 sessions. La taille des problèmes devient rapidement critique et les variables binaires, nombreuses, font grossir à souhait l'arbre de branchement. Le nombre des stations de base potentielles, en particulier, est un facteur qui limite la taille des problèmes.

#### Couplage stations de base-liens câblés et impact du nombre de stations de base potentielles

Quand la charge de trafic augmente, le couplage entre les liens câblés et les stations de base augmente. Ainsi, le nombre de sessions peut devenir suffisamment important pour permettre à la procédure d'améliorer le coût final en ajoutant une station de base tout en optimisant les capacités des liens câblés. Dans ce cas, les stations de base, en sur-nombre, seront sous-utilisées. Suivant les objectifs du dimensionnement, il peut être préférable de privilégier les stations de base. Cela demandera alors une approche qualitative pour jauger le coût des stations de base et des liens câblés de la fonction objectif compte tenu de la charge de trafic testée.

La multiplicité des stations de base potentielles permet d'affiner les positions des stations de base activées. Elle offre plus de possibilités à la procédure et multiplie ainsi les chances de réduire le nombre de stations de base sélectionnées, mais ce, sans véritable garantie. Il faut

donc trouver un nombre de stations de base potentielles raisonnable qui fasse le compromis entre la taille du problème et la possibilité de réduire au plus les stations de base activées.

### Optimalité des positions

Dans notre modèle, l'optimalité des positions des stations de base est liée à leur nombre. Ainsi, plus le nombre de stations de base est faible pour une instance de trafic donnée, plus les surfaces de couverture sont grandes. Les positions des stations de base sont donc meilleures pour cet aspect. Il est possible de trouver plusieurs combinaisons de stations de base avec le même coût pour le dimensionnement global. Pour privilégier une certaine combinaison, il faudrait définir et intégrer dans le modèle un critère d'optimalité des positions comme, par exemple, une répartition équitable du trafic sur les stations de base. Ce critère n'est valable que pour une zone de trafic homogène.

### Soft handover

Le modèle affine l'évaluation des ressources radio en tenant compte des usagers en *soft handover*. En effet, un usager pris en charge par plusieurs stations de base nécessite moins de puissance sur les liaisons radio établies, mais d'un autre côté, deux stations de base doivent être activées. Actuellement, le modèle ne force pas la procédure à prendre un usager mobile en *soft handover* quand ce dernier se trouve sur une zone commune de recouvrement entre deux stations de base activées. De plus, un usager mobile peut être pris en *soft handover* par deux stations de base très éloignées malgré d'autres stations de base plus proches. La répartition des sessions en *soft handover* obtenue pour une instance donnée n'est donc pas forcément réaliste, il manque des critères de distance plus élaborés que ceux déjà inclus dans notre modèle (ensemble  $L_s^{\text{BS-SH}}$ ). Le *soft handover* permet cependant à la procédure de minimiser au mieux les ressources quand le profil de trafic est tel que les stations de base activées sont en limite de charge maximale.

### Dimensionnement des réseaux 3G

La dernière partie du chapitre traite du modèle en tant qu'outil de dimensionnement. Les profils de trafic des instances testées décrivent le trafic sur une fenêtre de temps d'une minute. A nouveau, pour obtenir un problème de taille raisonnable, il a fallu faire un

compromis entre la durée de la planification et le nombre de sessions. Le générateur de trafic répartit les requêtes sur la durée de planification, mais les durées des connexions ne sont pas ramenées à la durée de planification, elles conservent des valeurs réalistes. Ainsi, un grand nombre de sessions qui durent plus d'une minute occuperont toutes les périodes suivantes à leur requête. De cette façon, on obtient une véritable photo du trafic écoulé sur une minute dans un réseau (figure 5.4). Nous avons donc privilégié le nombre de sessions plutôt que le nombre de périodes car il est plus réaliste de voir 200 sessions arriver sur une période courte que 100 sessions arriver sur période plus longue. Les dimensionnements obtenus nous ont permis d'observer qu'à terme, les liaisons radio descendantes seront le facteur limitant pour les capacités radio. De plus, les ressources devront être nombreuses pour supporter les nouveaux profils de trafic, en particulier les rayons de couverture des cellules seront réduits à quelques dizaines de mètres (pico-cellules), ce qui demandera un nombre important de stations de base à déployer.

## CONCLUSION

Dans ce mémoire, notre étude a porté sur la définition d'un modèle d'optimisation pour le dimensionnement des réseaux mobiles troisième génération. Plusieurs décisions ont été prises afin de modéliser le réseau et pour élaborer une stratégie de dimensionnement. Le premier défi pour la modélisation du réseau est de tenir compte des différents mécanismes qui permettent de garantir et de fournir une qualité de service adaptée au type d'application, concept clé pour les profils de trafic multimédia. A ce titre, nos choix se sont portés, d'une part, sur la modélisation d'un contrôle d'admission pour simuler les rejets de requêtes et pour garantir la qualité de service, et, d'autre part, sur la politique WFQ d'ordonnement des paquets pour fournir la qualité de service avec un partage des ressources adéquat. Notre modèle intègre la différenciation des services en traitant les usagers individuellement, c'est du service intégré. Cet aspect est important pour étendre les garanties de service personnalisées à la partie radio. Le contrôle d'admission modélisé s'appuie, entre autres, sur des formules de capacité radio adaptées au trafic multi-services. Le modèle intègre aussi le *soft handover* pour affiner l'évaluation des ressources radio nécessaires.

Le modèle proposé dans ce mémoire est une première étape vers un outil de dimensionnement global qui couple le coût des stations de base au coût des liens câblés. La stratégie de dimensionnement consiste à partir d'une configuration initiale de stations de base potentielles; la procédure d'optimisation active alors certaines stations de base et évalue les capacités optimales des liens câblés qui permettent de respecter les exigences du profil de trafic au moindre coût.

La démarche que nous avons suivie est la suivante.

- La compréhension des enjeux des futurs réseaux 3G a permis d'orienter la modélisation du réseau et du trafic pour rendre l'outil de dimensionnement adapté à cette nouvelle technologie. En particulier, la modélisation du réseau s'articule autour de deux parties: le réseau de cœur et les réseaux d'accès radio. Le trafic généré se base sur un profil multimédia et la qualité de service est assurée de bout en bout.
- Les objectifs du modèle sont déterminés par la stratégie de dimensionnement qui englobe à la fois les ressources radio et les ressources câblées. Le dimensionnement



final est tributaire de l'algorithme de contrôle d'admission utilisé par le modèle.

- L'étape suivante a consisté en la formulation mathématique du modèle sous forme d'un problème d'optimisation linéaire mixte.
- L'implémentation en C++ est organisée en deux parties : la génération des jeux de données et la génération du problème d'optimisation. Une fois le problème créé, le logiciel d'optimisation CPLEX-MIP procède à la résolution. L'accent a été mis sur la modélisation et l'obtention de résultats préliminaires, CPLEX-MIP est donc utilisé comme une pseudo boîte noire. Les stratégies d'exploration, de séparation et de branchement proposées par le logiciel sont testées sur plusieurs études de cas pour trouver les plus adaptées. De plus, sans aller jusqu'à développer un algorithme de séparation et d'évaluation propre, deux stratégies d'affectation de priorités sur les variables ont été élaborées.
- Les premiers résultats ont permis de valider le modèle en étudiant son comportement sur différents jeux de données. Puis, l'outil de dimensionnement obtenu est utilisé pour planifier un réseau test avec des profils de trafic différents.

Les résultats que nous avons obtenus permettent plusieurs conclusions.

- Un des algorithmes d'affectation de priorité sur les variables ainsi qu'une stratégie de branchement de CPLEX\_MIP se sont démarquées en étant plus efficaces sur l'ensemble des études de cas testées.
- Les temps de résolution des problèmes deviennent rapidement très grands à cause du nombre de variables binaires. Actuellement, des compromis pour réduire les jeux de données sont nécessaires. Il faut estimer de façon réaliste le rapport entre la durée de planification et le nombre de sessions; le nombre de stations de base potentielles est aussi un facteur important : il augmente les chances de réduire le nombre de stations de base activées mais sans véritable garantie et au prix d'une augmentation du nombre de variables binaires.
- Suivant le profil de trafic, l'outil de dimensionnement fourni par notre modèle privilégie, dans une certaine mesure, les ressources radio ou les capacités des liens câblés. En privilégiant les ressources radio, la procédure d'optimisation aura avant tout minimisé le nombre de stations de base en maximisant leurs rayons de couverture. A

l'inverse, ce nombre pourra être plus élevé dans le dimensionnement final si la priorité est mise sur les capacités des liens câblés.

- Le dimensionnement optimal est associé à un contrôle d'admission anticipatif idéal. Les critères de refus se basent sur toutes les requêtes antérieures et postérieures, sur les durées des sessions et sur les routes afin que le contrôle d'admission évalue leurs impacts sur l'utilisation des ressources.
- Le modèle respecte les exigences de qualité de service et les niveaux de service selon les applications. Ainsi, l'utilisation des ressources est fidèle à une répartition multiservices. Les résultats obtenus permettent de se rendre compte de l'importance du dimensionnement pour ces futurs réseaux. Il faudra déployer des ressources importantes pour supporter les nouveaux profils de trafic très exigeants. En particulier, les tailles des cellules devraient être réduites à la dizaine de mètres pour assurer la couverture des applications comme la navigation Internet. Ces applications ont de fortes exigences de débit dans le sens descendant, ce qui rendra critique les niveaux de capacités occupées sur les liaisons radio descendantes.

Comme tout travail de modélisation, il a fallu, dans le cadre de notre étude, accepter certaines limites et restrictions au modèle.

- Dans un premier temps, la taille des problèmes est tel que, sans branchement approprié, la procédure d'optimisation ne permet pas de résoudre de façon satisfaisante des profils de trafic importants. Il serait intéressant de développer une stratégie de branchement qui exploiterait les relations entre les variables. Une autre possibilité pour améliorer les temps de résolution est de partir d'une solution initiale entière trouvée de façon heuristique. Cette approche est étudiée dans <sup>[33]</sup> où la méthode Tabou est utilisée pour obtenir une solution entière réalisable à notre problème d'optimisation.
- La répartition du trafic est supposée homogène sur l'aire géographique à dimensionner : cette hypothèse est nécessaire pour les formules de capacité radio multiservices. Elle devient cependant restrictive si la surface à dimensionner est étendue et si la durée de planification s'étale sur une journée. En effet, dans un centre urbain, les différents pics de trafic observés sur une journée ne sont pas concentrés au même endroit ; pour un meilleur dimensionnement il faudrait généraliser les formules de capacité radio au

cas inhomogène.

- Le séquençement temporel des sessions conduit à une surestimation de leurs durées. De plus, les durées sont connues par avance et donc indépendantes des débits alloués. Cela entraîne une imprécision sur la durée d'utilisation des ressources par les applications non temps réel. Une possibilité à envisager serait de fonctionner avec des scénarios associant la durée au débit alloué.
- Le contrôle d'admission du modèle n'a pas de contrainte de causalité, il s'éloigne ainsi des véritables performances des contrôles d'admission réels. Le contrôle d'admission anticipatif a plus de marge de manœuvre pour réduire le coût du dimensionnement que les contrôles d'admission réels. Le dimensionnement obtenu pourrait ainsi servir de référence pour comparer l'efficacité de différents contrôles d'admission implémentables. Dans le cadre d'une autre étude, une procédure itérative basée sur le principe du modèle a été développée dans le but de forcer le contrôle d'admission à ne tenir compte que des requêtes courantes, ce contrôle d'admission est dit causal. Cela a permis de mener une étude comparative<sup>[34]</sup> des dimensionnements obtenus avec les deux contrôles d'admission.
- Le modèle ne permet pas une répartition complètement réaliste des usagers mobiles en *soft* handover sur les stations de base, des contraintes spatiales plus élaborées pourraient être ajoutées au modèle pour améliorer cet aspect.
- Une autre extension à ce modèle est en cours d'études et concerne le multi-routage.
- Un autre aspect à développer serait la mobilité des usagers, comment en tenir compte pour les routes des connexions qui ne seraient plus statiques et comment évaluer les ressources radio pour ces usagers en déplacement?

## RÉFÉRENCES

- [1] G Bertini, "Subscriber Traffic Profile For CDMA2000 1XRTT," Tech. Rep. PB1, Ericsson Documentation, août 2000.
- [2] 3GPP. <http://www.3gpp.org>.
- [3] 3GPP2. <http://www.3gpp2.org>.
- [4] H Holma et Antti Toskala, *WCDMA for UMTS*. Wiley, 2000.
- [5] D Wisely et P Eardley et L Burness, *IP for 3G. Networking Technologies for Mobile Communications*. John Wiley et Sons LTD, 2002.
- [6] IETF. <http://www.ietf.org/rfc.html>.
- [7] 3GPP , "3rd Generation Partnership Project. Technical Groupe Services and System Aspects; QoS Concepts and Architecture (Release 1999)," Tech. Rep. TS 23.107 v3.2.0, <http://www.3gpp.org>, 2000.
- [8] R. Cox, "Three Speech Coders from the ITU Cover a Range of Applications ," *IEEE Communications Magazine*, novembre 1997.
- [9] J Meggers et M Jägers et R.S Barathy, "Multimedia Communications in Future Packet-Switched Wireless Communication Networks," tech. rep., Ericsson Documentation, juillet 1998.
- [10] R Guérin et H Ahmadi et M Naghshineh, "Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, septembre 1991.
- [11] J Walrand et P Varaiya, *High Performance Communication Networks*. Morgan Kaufmann Publisher, 2000.
- [12] S Jamin et P.B Danzig et S Shenker et L Zhang, "A Measurement-based Admission Control Algorithm for Integrated Services Packet Networks ," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 56–70, février 1997.
- [13] D Levine et M Naghshineh et I Akildyz, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 1–12, février 1997.

- [14] K Lee, "Supporting Mobile Multimedia in Integrated Services Networks," *Wireless Networks, ACP Press*, vol. 2, août 1996.
- [15] D.C Verna et H Zhang et D Ferrari, "Delay Jitter for Real-Time Communication in a Packet Switching Network," *Proceedings of TriComm'91*, 1991.
- [16] H Zhang, "Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks," *IEEE Proceedings*, vol. 83, octobre 1995.
- [17] H.D Sherali et C.M Pendyala et T.S Rappaport, "Optimal Location of Transmitters for Micro-Cellular Radio Communication System Design," *IEEE Journal on Selected Areas in Communications*, vol. 14, mai 1996.
- [18] K Tutschku, "Demand-based Radio Network Planning of Cellular Mobile Communication Systems," *Proc IEEE Infocom'98*, pp. 1054–1061, Avril 1998.
- [19] C.Y Lee et H.G Kang, "Cell Planning Capacity Expansion in Mobile Communications: A Tabu Search Approach," *IEEE Transactions on Vehicular Technology*, vol. 49, septembre 2000.
- [20] D Mehdi et D Tipper, "Some Approaches to Solving a Multi-Hour Broadband Network Capacity Design Problem with Single-Path Routing," *Telecommunication Systems*, vol. 13, no. 2, pp. 269–291, 2000.
- [21] D Mehdi et S Guptan, "Network Dimensioning and Performance of Multiservice, Multirate loss networks with dynamic routing," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 944–957, décembre 1997.
- [22] A Demers et S keshav et S Shenker, "Analysis and Simulation of a Fair Queuing Algorithm," *Journal of Internetworking: Research and Experience*, vol. 1, pp. 3–26, septembre 1990.
- [23] A K Parekh et R G Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 344–357, juin 1993.
- [24] A K Parekh et R G Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple-Node Case," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 137–150, avril 1994.

- [25] Z.L. Zhang et D Towsley et J Kurose, "Statistical Analysis of Generalized Processor Sharing Scheduling Discipline," *Journal on Selected Areas in Communications.IEEE*, vol. 13, pp. 368–379, août 1995.
- [26] J Nagle, "On Packet Switches with Infinite Storage," *IEEE Transactions on Communication*, avril 1987.
- [27] S Madan et M Diez, "IS-2000 Forward Link Simulation. Performance Results," Tech. Rep. v6, Ericsson Documentation, 2000.
- [28] B Jaumard et R Malhamé et O Troeung et S Esposito et C Klam et C Voisin , "Deliverable 2: Call Admission Control Policies and Dimensioning of Multimedia 3G Networks-Simulation Plan," tech. rep., École Polytechnique de Montréal, avril 2002. Ericsson Contract #1, Deliverable #2, Revised.
- [29] W.C Lee, *Mobile Cellular Telecommunications Systems. p112-113*. McGraw-Hill, 1989.
- [30] B Lind, "Application Traffic Model For UMTS Services," Tech. Rep. PD4, Ericsson Documentation, 2000.
- [31] ILOG. <http://www.ILOG.com>.
- [32] ILOG, *ILOG CPLEX 7.0 Référence Manual*, 2000. .
- [33] A Fortin et N Hail et B Jaumard, "A Tabu Search Heuristic for the Dimensioning of 3G Multi-services Networks," *Soumis pour publication*, 2002.
- [34] B Jaumard et Y Solari et C Voisin, "Causal and Anticipative Models for the Dimensioning of 3G Multiservice Networks," *Soumis pour publication dans ICC 2003*, 2002.

## Annexe I

## Protocole de Répétition Sélective

## I.1 Principe

Le protocole de répétition sélective ou SRP est utilisé comme protocole de retransmission sur le lien radio quand la trame est perdue ou erronée.

L'expéditeur, qui peut être la station de base ou l'utilisateur mobile, envoie une série de trames. Si l'origine détecte une trame erronée alors elle envoie à l'origine un acquittement négatif (NACK) contenant le numéro de la trame erronée. Dès réception de l'acquittement négatif, l'origine arrête la transmission séquentielle des trames pour renvoyer la trame demandée puis elle reprend l'émission là où elle s'était arrêtée. La retransmission est sélective car l'expéditeur ne renvoie que la trame erronée. Avec des protocoles comme le Go Back N, l'expéditeur aurait recommencé la transmission depuis la trame erronée.

Avec le protocole à répétition sélective, les trames sont numérotées pour conserver les séquences dans le bon ordre. De plus, la taille de la mémoire tampon du destinataire doit être suffisamment grande pour conserver toutes les trames envoyées depuis la réception d'une trame erronée pour pouvoir réorganiser la séquence après la retransmission.

La figure I.1 montre l'arrivée séquentielle des trames quand une erreur se produit. L'erreur se produit sur la trame Tr(2), le destinataire conserve en mémoire les trames Tr(3) et Tr(4) jusqu'à réception de la trame Tr(2) pour pouvoir rétablir l'ordre de la séquence.

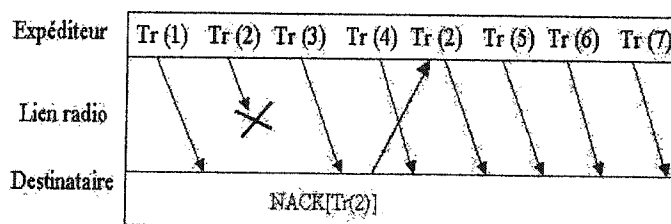


FIG. I.1 – *Ordre des retransmissions avec le protocole SRP*

## I.2 Délais moyens de transmission

Quand il y a retransmission, le délai de la trame est pénalisé par des délais supplémentaires qui sont les temps de retransmission de la trame mais aussi les temps de transmission des acquittements négatifs. On néglige les délais d'attente de la trame qui s'ajoutent quand une trame précédente doit être retransmise. La probabilité d'une erreur sur une trame est donnée par le taux d'erreur, ou  $FER$ , du RAB sélectionné pour la transmission.

Notons  $D_{ntl}$  le temps moyen de transmission d'une trame sur le lien radio et  $D_{to}$  le temps entre deux NACKs successifs. Après l'envoi du premier NACK, l'expéditeur attend  $D_{to}$  pour renvoyer un NACK si une erreur s'est à nouveau produite. Il fera de même pour les prochains NACKs.

Notons  $D_{trame}$  le délai moyen d'une trame en tenant compte des éventuelles retransmissions et  $D_{trame}^k$  de délai de la trame quand  $k$  retransmissions ont été nécessaires.

### – Sans retransmission

- Probabilité de ne pas avoir d'erreur  $P^0 = 1 - FER$ ,
- $D_{trame}^0 = D_{ntl}$ .

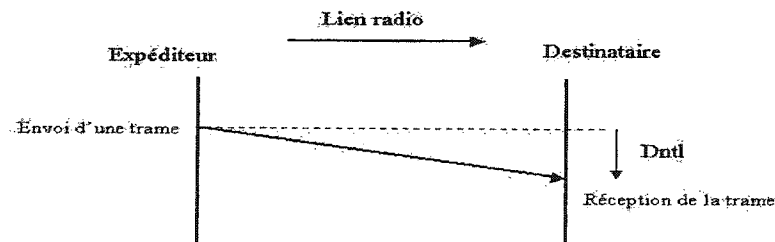


FIG. I.2 – Délai sans retransmission



– Une retransmission

- Probabilité d'avoir une erreur sur deux transmissions  $P^1 = FER \times (1 - FER)$ ,
- $D_{\text{trame}}^1 = 3 \times D_{\text{ntl}}$ .

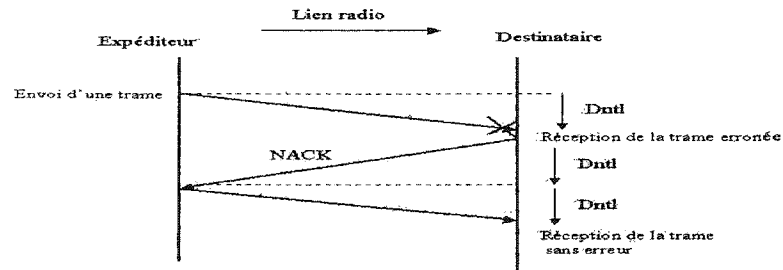


FIG. I.3 – Délai avec une retransmission

– Deux retransmissions

- Probabilité d'avoir deux erreurs sur trois transmissions  $P^2 = FER \times FER \times (1 - FER)$ ,
- $D_{\text{trame}}^2 = 3 \times D_{\text{ntl}} + D_{\text{to}}$ .

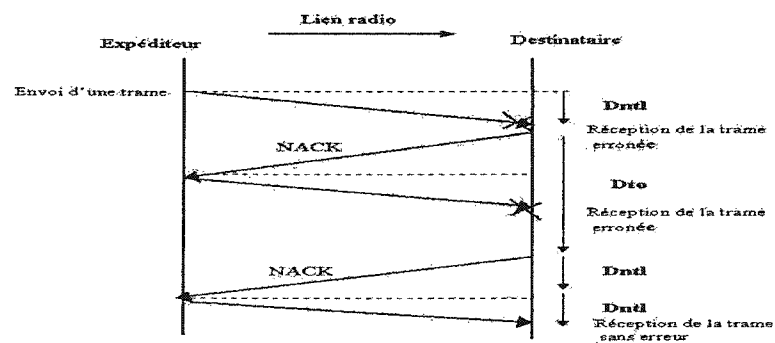


FIG. I.4 – Délai avec 2 retransmissions

– Trois retransmissions

- Probabilité d'avoir trois erreurs sur quatre transmissions  $P^3 = FER \times FER \times FER \times (1 - FER)$ ,
- $D_{\text{trame}}^3 = 3 \times D_{\text{ntl}} + 2 \times D_{\text{to}}$ .

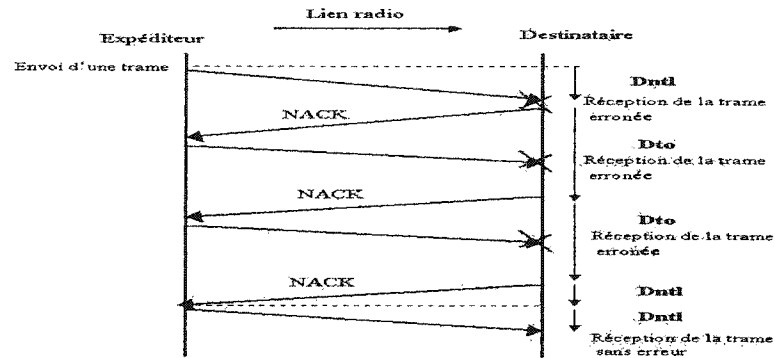


FIG. I.5 – Délai avec 3 retransmissions

En négligeant le cas où plus de trois retransmissions seraient nécessaires ( $FER$  faible), le temps moyen de transmission d'une trame s'écrit :

$$D_{\text{trame}} = P^0 \times D_{\text{trame}}^0 + P^1 \times D_{\text{trame}}^1 + P^2 \times D_{\text{trame}}^2 + P^3 \times D_{\text{trame}}^3$$

On obtient donc :

$$D_{\text{trame}} = (1 - FER) \times D_{\text{ntl}} + FER \times (1 - FER) \times (3 \times D_{\text{ntl}}) + FER^2 \times (1 - FER) \times (3 \times D_{\text{ntl}} + D_{\text{to}}) + FER^3 \times (1 - FER) \times (3 \times D_{\text{ntl}} + 2 \times D_{\text{to}}).$$

Avec  $D_{\text{ntl}} = 100\text{ms}$  et  $D_{\text{to}} = 2 \times D_{\text{ntl}} + 60 = 260\text{ms}$ , on obtient les délais suivant :

$FER=0.5 \%$	$D_{\text{trame}} = 101.01 \text{ ms}$
$FER=1 \%$	$D_{\text{trame}} = 102.03 \text{ ms}$
$FER=2 \%$	$D_{\text{trame}} = 104.11 \text{ ms}$
$FER=5 \%$	$D_{\text{trame}} = 110.68 \text{ ms}$
$FER=10 \%$	$D_{\text{trame}} = 122.78 \text{ ms}$

## Annexe II

### Paramètres du générateur de trafic

Les attributs des sessions sont générés selon plusieurs lois de probabilités, chacune caractérisée par un ou des paramètres qui déterminent la valeur moyenne et la variance. Cette annexe récapitule les valeurs numériques de ces paramètres utilisées dans le générateur de trafic.

#### II.1 Paramètres pour la Voix

TAB. II.1 – Valeurs des paramètres pour l'application Voix

	Distribution	Paramètre	Valeur du paramètre
Temps d'arrivée	poisson	$\lambda$	$\frac{NBR\_SESSIONS\_VOIX}{DUREE\_PLANIFICATION}$
Temps d'inter-arrivée	exponentielle	$\lambda$	$\frac{NBR\_SESSIONS\_VOIX}{DUREE\_PLANIFICATION}$
Durée	exponentielle	DUREE_MOY	31 sec
Priorité	bernouilli	PROBABILITE_GOLD	0.7

#### II.2 Paramètres pour la Vidéo conférence

TAB. II.2 – Valeurs des paramètres pour l'application Vidéo conférence

	Distribution	Paramètre	Valeur du paramètre
Temps d'inter-arrivée	lognormal	moyenne= $\frac{1}{\lambda}$ variance= $\frac{1}{\lambda^2}$	$\lambda = \frac{NBR\_SESSIONS\_VIDEOCONF}{DUREE\_PLANIFICATION}$
Durée	exponentielle	DUREE_MOY	180 sec
Priorité	bernouilli	PROBABILITE_GOLD	0.7
Type	bernouilli	PROBABILITE_HIGH	0.3

## II.3 Paramètres pour le Flux vidéo

TAB. II.3 – Valeurs des paramètres pour l'application Flux vidéo

	Distribution	Paramètre	Valeur du paramètre
Temps d'arrivée	poisson	$\lambda$	$\frac{NBR\_SESSIONS\_FLUX\_VIDEO}{DUREE\_PLANIFICATION}$
Temps d'inter-arrivée	exponential	$\lambda$	$\frac{NBR\_SESSIONS\_FLUX\_VIDEO}{DUREE\_PLANIFICATION}$
Durée	exponentielle	DUREE_MOY	180 sec
Priorité	bernouilli	PROBABILITE_GOLD	0.7
Type	bernouilli	PROBABILITE_HIGH	0.3

## II.4 Paramètres pour la Navigation Internet

TAB. II.4 – Valeurs des paramètres pour l'application Navigation Internet

	Distribution	Paramètre	Valeur du paramètre
Temps d'inter-arrivée	lognormal	mean= $\frac{1}{\lambda}$ variance= $\frac{1}{\lambda^2}$	$\lambda = \frac{NBR\_SESSIONS\_NAVIGATION}{DUREE\_PLANIFICATION}$
Durée	exponentielle	DUREE_MOY	250 sec
Priorité	bernouilli	PROBABILITE_GOLD	0.5
Type	bernouilli	PROBABILITE_HIGH	0.3

## II.5 Paramètres pour le Courrier électronique

TAB. II.5 – Valeurs des paramètres pour l'application Courrier électronique

	Distribution	Paramètre	Valeur du paramètre
Temps d'arrivée	poisson	$\lambda$	$\frac{NBR\_SESSIONS\_COURRIER}{DUREE\_PLANIFICATION}$
Temps d'inter-arrivée	exponential	$\lambda$	$\frac{NBR\_SESSIONS\_COURRIER}{DUREE\_PLANIFICATION}$
Durée	exponentielle	DUREE_MOY	10 sec
Priorité	bernouilli	PROBABILITE_GOLD	0.5
Direction	bernouilli	PROBABILITE_DL	0.5

### Annexe III

#### Exemple de calcul des différents délais liés au flot descendant d'une session VBR

Les calculs se basent sur les valeurs numériques des tableaux 5.11 et 5.12. On suppose, par la suite, que le nombre de liens câblés sur la route du flot descendant de la session est fixé à 5 (le lien entre le RNC et la station de base est inclus). Rappelons les formules des différents délais qui interviennent dans la contrainte (4.20) :

$$(4.15) \quad D_{f_s^{DL-WL}}^{max-trans} = m_{f_s^{DL}} \times \frac{L_{f_s^{DL-WL}}}{\mathcal{R}(r_t^{DL})},$$

$$(3.26) \quad D_{f_s^{DL-WL}}^{propag} = m_{f_s^{DL}} \times Temps\_propagation/liens,$$

$$(3.28) \quad D_{f_s^{DL-WL}}^{trait} = m_{f_s^{DL}} \times Temps\_traitement/routeur,$$

$$(3.34) \quad D_{f_s^{DL-RL}}^r = N_{r_t^{DL}} \times D_{r_{FER}}^{transmission-radio},$$

$$(3.32) \quad N_{r_t^{DL}} = \left\lfloor \frac{8 \times L_{f_s^{DL-WL}}}{Duree\_trame \times r_t^{DL}} \right\rfloor,$$

avec :

- $m_{f_s^{DL}} = 5$ ,
- $Duree\_trame = 20 \text{ ms}$ ,
- Les valeurs  $\mathcal{R}(r_t^{DL})$  sont données dans le tableau 3.3,
- L'ensemble des RABs,  $T_s^{DL}$ , est détaillé dans le tableau 5.9 pour la navigation Internet High et Low et dans le tableau 5.10 pour le courrier électronique,
- Les valeurs de  $D_{r_{FER}}^{transmission-radio}$  sont données dans le tableau 3.5.

### III.1 Calcul des délais pour les sessions de navigation Internet

TAB. III.1 – *Délais pour une session de navigation Internet*

Navigation Internet Low				Navigation Internet High			
$D_s^{D-MAX}$		5 s		$D_s^{D-MAX}$		2 s	
$D_{f_{DL-WL}}^{propag}$		0.05 s		$D_{f_{DL-WL}}^{propag}$		0.05 s	
$D_{f_{DL-WL}}^{trait}$		0.10 s		$D_{f_{DL-WL}}^{trait}$		0.10 s	
$T_s^{DL}$	$D_{f_{DL-WL}}^{max-trans,rt}$	$N_{rt_{DL}}$	$D_{f_{DL-RL}}^r$	$T_s^{DL}$	$D_{f_{DL-WL}}^{max-trans}$	$N_{rt_{DL}}$	$D_{f_{DL-RL}}^r$
r1 [76.8 kbps; 5%]	0.4 s	5	0.554 s	r1 [307.2 kbps; 5%]	0.1 s	1	0.111 s
r2 [76.8 kbps; 10%]	0.4 s	5	0.614 s	r2 [307.2 kbps; 10%]	0.1 s	1	0.123 s
r3 [38.4 kbps; 5%]	0.8 s	9	0.997 s	r3 [153.6 kbps; 5%]	0.2 s	2	0.222 s
r4 [38.4 kbps; 10%]	0.8 s	9	1.105 s	r4 [153.6 kbps; 10%]	0.2 s	2	0.246 s

### III.2 Calcul des délais pour les sessions de courrier électronique dans le sens descendant

TAB. III.2 – *Délais pour une session de courrier électronique*

Courrier électronique (DL)			
$D_s^{D-MAX}$		10 s	
$D_{f_{DL-WL}}^{propag}$		0.10 s	
$D_{f_{DL-WL}}^{trait}$		0.05 s	
$T_s^{DL}$	$D_{f_{DL-WL}}^{max-trans,rt}$	$N_{rt_{DL}}$	$D_{f_{DL-RL}}^r$
r1 [153.6 kbps; 5%]	0.2 s	2	0.222 s
r2 [153.6 kbps; 10%]	0.2 s	2	0.246 s
r3 [76.8 kbps; 5%]	0.4 s	5	0.554 s
r4 [76.8 kbps; 10%]	0.4 s	5	0.614 s
r5 [38.4 kbps; 5%]	0.8 s	9	0.997 s
r6 [38.4 kbps; 10%]	0.8 s	9	1.105 s
r7 [19.2 kbps; 5%]	1.66 s	17	1.882 s
r8 [19.2 kbps; 10%]	1.66 s	17	2.088 s

### III.3 Calcul des valeurs minimales de débit pour le flot descendant imposées par la contrainte de délai.

La contrainte (4.20) impose un débit minimal au flot descendant sur les liens câblés. Rappelons la forme de la contrainte :

$$x_{f_{DL-WL}}^i \geq \sum_{r \in T_s^{DL}} \max \left( \mathcal{R}(r_t), \frac{m_{f_{DL}} L_{f_{DL}}^{max} + \sigma_s}{(D_s^{DL-MAX} - D_{f_{DL-WL}}^{max-trans,rt} - D_{f_{DL-WL}}^{propag} - D_{f_{DL-WL}}^{trait} - D_{f_{DL-RL}}^r)} \right) \alpha_{ir} f_{DL-RL}^r$$

En injectant les valeurs des délais calculées dans les deux sections précédentes, on obtient les deux valeurs possibles pour le débit minimal.

TAB. III.3 – Débits minimum pour le flot  $f_s^{\text{DL-WL}}$  d'une session de navigation Internet

Navigation internet Low		
$T_s^{\text{DL}}$	$\mathcal{R}(r_t)$	$\frac{m_{f_{\text{DL}}} L_{f_{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-WL}}}^r)}$
r1 [76.8 kbps , 5%]	40 kbps	17.99 kbps
r2 [76.8 kbps , 10%]	40 kbps	18.27 kbps
r3 [38.4 kbps , 5%]	20 kbps	22.95 kbps
r4 [38.4 kbps , 10%]	20 kbps	23.78 kbps
Navigation internet High		
$T_s^{\text{DL}}$	$\mathcal{R}(r_t)$	$\frac{m_{f_{\text{DL}}} L_{f_{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-WL}}}^r)}$
r1 [307.2 kbps , 5%]	160 kbps	42.76 kbps
r2 [307.2 kbps , 10%]	160 kbps	43.07 kbps
r3 [153.6 kbps , 5%]	80 kbps	49.08 kbps
r4 [153.6 kbps , 10%]	80 kbps	49.91 kbps

TAB. III.4 – Débits minimum pour le flot  $f_s^{\text{DL-WL}}$  d'une session de courrier électronique

Courrier électronique (sens DL)		
$T_s^{\text{DL}}$	$\mathcal{R}(r_t)$	$\frac{m_{f_{\text{DL}}} L_{f_{\text{DL}}}^{\text{max}} + \sigma_s}{(D_s^{\text{DL-MAX}} - D_{f_s^{\text{DL-WL}}}^{\text{max-trans}, r_t} - D_{f_s^{\text{DL-WL}}}^{\text{propag}} - D_{f_s^{\text{DL-WL}}}^{\text{trait}} - D_{f_s^{\text{DL-WL}}}^r)}$
r1 [153.6 kbps , 5%]	80 kbps	7.43 kbps
r2 [153.6 kbps , 10%]	80 kbps	7.45 kbps
r3 [76.8 kbps , 5%]	40 kbps	7.89 kbps
r4 [76.8 kbps , 10%]	40 kbps	7.93 kbps
r5 [38.4 kbps , 5%]	20 kbps	8.70 kbps
r6 [38.4 kbps , 10%]	20 kbps	8.82 kbps
r7 [19.2 kbps , 5%]	10 kbps	11.11 kbps
r8 [19.2 kbps , 10%]	10 kbps	11.48 kbps