

UNIVERSITÉ DE MONTRÉAL

CONTRÔLE D'ADMISSION AVEC MESURES
POUR UNE MEILLEURE GESTION DES RESSOURCES
DANS LES RÉSEAUX DE TROISIÈME GÉNÉRATION

SANDRINE ESPOSITO
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLOME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES (M.Sc.A.)
(TÉLÉCOMMUNICATIONS)
AVRIL 2003



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.

0-612-81544-7

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

CONTRÔLE D'ADMISSION AVEC MESURES
POUR UNE MEILLEURE GESTION DES RESSOURCES
DANS LES RÉSEAUX DE TROISIÈME GÉNÉRATION

présenté par: ESPOSITO Sandrine

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

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

M. PIERRE Samuel, Ph.D., président

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

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

M. MABILLEAU Philippe, Ph.D., membre

À mes parents, à mes frères

REMERCIEMENTS

Je tiens à remercier ma directrice de recherche, Brigitte Jaumard, dont le suivi et l'aide financière prodigués ont permis l'aboutissement de cette maîtrise. Ce travail a été mené en collaboration avec *Ericsson Research Canada*, et j'adresse mes remerciements à Francis Lupien et Daniel Dufour qui ont su guider nos pas dans ce projet à connotation industrielle. De plus, j'adresse mes remerciements au professeur Roland Malhamé qui nous a aidés durant les premiers mois de la maîtrise.

Je remercie également Monsieur Richard Hurteau pour son aide, ainsi que Daniel Charbonneau pour sa disponibilité et son suivi sur Opnet.

Je tiens à remercier tout particulièrement ma famille pour son soutien de tous les instants, mon papa et ma maman qui ont toujours été là pour moi, et mes frères que j'adore!

Un grand merci aussi à mon Thomas, qui a accompagné ma vie ici, et je ne peux oublier Cathy, à la fois amie de coeur et partenaire de travail!

Mes remerciements vont finalement à Supélec (l'*École Supérieure d'Electricité*), mon école d'origine, qui m'a permis de partir à l'étranger et vivre cette aventure.

RÉSUMÉ

Le succès commercial de la téléphonie mobile, le nombre croissant d'utilisateurs de l'Internet et l'émergence de nouvelles applications multimédias sont les raisons essentielles de l'évolution vers les réseaux de troisième génération. La 3G verra son architecture entièrement fondée sur IP afin de fournir une large variété de services, parmi lesquels se trouvent des applications multimédias temps réel et non temps réel possédant des exigences de qualité de service différenciées. Les réseaux cellulaires actuels ne pouvant fournir qu'un service "meilleur effort", la garantie de QoS et la gestion efficace des ressources sont des aspects essentiels dans la conception des réseaux 3G.

Dans ce mémoire, nous proposons d'utiliser un algorithme de contrôle d'admission basé sur les mesures intégrant une politique de gestion des ressources adaptative afin de supporter les contraintes de QoS, notamment de délai, des applications temps réel et non temps réel, tout en atteignant un taux d'utilisation des ressources satisfaisant. Se fondant sur l'estimation de la capacité CDMA multiservices dans le réseau d'accès radio, et sur les propriétés de la discipline d'attente équitable proportionnelle dans le réseau cœur, nous établissons une architecture globale de contrôle et de décision nouvelle.

Un simulateur prototype est développé sous le logiciel de modélisation et de simulation de réseaux OPNET afin d'étudier les performances du système pour des répartitions typiques de trafic dans une cellule radio. Les caractéristiques spécifiques de l'algorithme de contrôle d'admission basé sur les mesures sont analysées puis ses performances comparées avec un algorithme basé sur des descripteurs de trafic à priori. L'étude permet ainsi de valider l'architecture proposée qui peut alors servir de base pour la conception et le dimensionnement des futurs réseaux 3G.

ABSTRACT

The commercial success of mobile telephony, increasing number of Internet users, and the emergence of new multimedia applications are the major reasons behind third-generation evolution. 3G will operate on an overall IP-based framework to provide a wide variety of services, among which multimedia real time and non real time applications have differentiated quality of service requirements. Current cellular networks being able to provide only a “best effort” service, guarantee of QoS and effective resources management are thus essential aspects in the design of 3G networks.

In this thesis, we propose to use a measurement based admission control algorithm integrating an adaptive resources management policy in order to support QoS constraints, like delay constraint, of real time and non real time applications, while reaching a satisfying utilization of resources. Basing ourselves on the estimate of multiservices CDMA capacity in the radio access network, and on the properties of the weighted fair queueing discipline in the core network, we establish a new global architecture of control and decision.

A prototype simulator is developed with the OPNET network modeling and simulation software to study the system performances for typical traffic distributions within a radio cell. Specific characteristics of the measurement based admission control algorithm are analyzed, then its performances are compared with a parameter-based admission algorithm. The study thus validates the proposed architecture, which can then be used as a basis for the design and dimensioning of future 3G networks.

TABLE DES MATIÈRES

DÉDICACE	iv
REMERCIEMENTS	v
RÉSUMÉ	vi
ABSTRACT	vii
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	xii
LISTE DES FIGURES	xiv
LISTE DES NOTATIONS ET DES SYMBOLES	xvii
LISTE DES ANNEXES	xx
INTRODUCTION	xxi
0.1 Motivations	xxi
0.2 Organisation du mémoire	xxii
0.3 Contributions	xxiii
CHAPITRE 1: ARCHITECTURE ET SERVICES DES RÉSEAUX 3G	1
1.1 Présentation générale des réseaux de troisième génération	1
1.1.1 Historique des systèmes cellulaires	1
1.1.2 Systèmes cellulaires de troisième génération	2
1.2 Le réseau cœur CDMA2000	6
1.2.1 Architecture	6
1.2.2 Mobilité IP	8
1.3 Le réseau d'accès radio	9

1.3.1	La méthode d'accès radio CDMA	9
1.3.2	Contrôle de puissance en CDMA	15
1.3.3	Macrodiversité en CDMA	16
1.3.4	Les canaux physiques CDMA2000	18
1.4	La qualité de service	20
1.4.1	Critères de qualité de service	20
1.4.2	Classes de service différenciées et priorités d'utilisateur	21
1.4.3	Modèle à intégration de services pour le réseau cœur	24
1.4.4	Implémenter la qualité de service	27
1.4.5	Garantie de délai	34
1.5	Le contrôle d'admission dans le système CDMA du réseau radio	39
1.6	Le contrôle d'admission dans le réseau cœur	41
1.6.1	Contrôle d'admission pour le service garanti	42
1.6.2	Contrôle d'admission pour le service statistique	43
1.6.3	Contrôle d'admission pour le service prédictif	43
1.7	Conclusion	47
CHAPITRE 2: EVALUATION DE LA CAPACITÉ RADIO CDMA2000		48
2.1	Estimations de la capacité CDMA dans la littérature	49
2.2	Modèle d'évaluation de la capacité sur le lien montant	49
2.2.1	Hypothèses	50
2.2.2	Définition des paramètres	51
2.2.3	Mise en équations du modèle sans <i>soft handover</i>	51
2.2.4	Equations avec <i>soft handover</i>	54
2.3	Modèle d'évaluation de la capacité sur le lien descendant	56
2.3.1	Hypothèses	56
2.3.2	Définition des paramètres	57
2.3.3	Mise en équations du modèle sans <i>soft handover</i>	58
2.3.4	Equations avec <i>soft handover</i>	61
2.4	Validation des équations de capacité	64

2.4.1	Valeurs numériques des paramètres utilisés	64
2.4.2	Analyse des équations	64
2.4.3	Influence du rayon de la cellule dans le sens descendant	67
2.4.4	Influence du type de trafic	68
2.5	Conclusion	71
CHAPITRE 3: CONTRÔLE D'ADMISSION POUR LES RÉSEAUX 3G		73
3.1	Contexte d'admission	73
3.1.1	Modélisation du réseau	73
3.1.2	Filtrage du trafic et discipline d'ordonnancement dans le réseau cœur	74
3.1.3	Ressources sur l'interface radio	75
3.1.4	Composantes du délai bout en bout	76
3.1.5	Garantie de délai et de bande passante	81
3.1.6	Spécification de QdS	82
3.2	Contrôle d'admission pour du service garanti	85
3.2.1	Notations	85
3.2.2	Vérification des ressources disponibles	87
3.2.3	Réservation et allocation des ressources	90
3.2.4	Renégociation de QdS	90
3.3	Contrôle d'admission basé sur les mesures	93
3.3.1	Le processus de mesure	94
3.3.2	Spécification de QdS	96
3.3.3	Vérification des ressources disponibles	97
3.3.4	Réservation et allocation des ressources	97
3.3.5	Libération de ressources	98
3.3.6	Algorithme d'admission MBCAC	101
3.4	Conclusion	101
CHAPITRE 4: RÉSULTATS		104
4.1	Description du simulateur	105

4.1.1	Réseau simulé	105
4.1.2	Modèles de trafic	107
4.1.3	Paramètres de configuration annexes	111
4.2	Les différents cas d'étude	113
4.2.1	Scénario (A)	113
4.2.2	Scénario (B)	114
4.3	Validation de performances	117
4.4	Etudes et performances de l'algorithme MBCAC en fonction des paramètres de mesure	119
4.4.1	Phénomène de blocage	119
4.4.2	Taux d'occupation à la station de base	120
4.4.3	Garantie de QdS	123
4.5	Etude comparative avec la procédure CAC1	126
4.5.1	Phénomène de blocage	127
4.5.2	Taux d'occupation à la station de base	131
4.5.3	Garantie de QdS	133
4.6	Aide au dimensionnement	134
4.7	Conclusion	135
CONCLUSION		136
RÉFÉRENCES		140

LISTE DES TABLEAUX

1.1	Caratéristiques d'accès radio des principaux systèmes cellulaires de deuxième génération. VD: voie descendante, VM: voie montante, QPSK: quadrature phase shift keying, GMSK: gaussian minimum shift keying	2
1.2	Caratéristiques des technologies d'accès radio de CDMA2000 et de l'UMTS (VD:voie descendante, VM: voie montante).	5
1.3	Les 4 classes de service définies par 3GPP	23
1.4	Hierarchisation relative des sessions suivant le type d'application et la priorité de l'utilisateur	24
1.5	Comparaison des différents MBCAC	46
2.1	Les équations de capacité radio dans la littérature	50
2.2	Paramètres des équations de capacité radio	64
2.3	RABS par types d'application	69
2.4	Les différentes répartitions de trafic	70
3.1	Exigences de QoS d'une session faisant une requête au réseau	83
3.2	Descripteurs de trafic	84
4.1	Sessions d'utilisateurs	106
4.2	Standards audios recommandés par l'UIT	108
4.3	Standards vidéos principaux	108
4.4	Paramètres de configuration des applications types dans le simulateur . .	109
4.5	Paramètres des exigences de QoS des applications dans le simulateur . . .	109
4.6	Paramètres pour le réseau cœur	112
4.7	Configuration des paramètres radios	112
4.8	Capacité utilisée par chaque type de session	113
4.9	Paramètres de configuration des profils de trafic pour le scénario (A) . . .	114
4.10	Paramètres de configuration des profils de trafic pour le scénario (B) . . .	116

4.11 Taux d'utilisation comparés pour les scénarios A et B en fonction des paramètres de mesure	123
4.12 Capacité du système dans le sens descendant	127
4.13 Taux d'utilisation comparés pour les scénarios A et B sous les procédures CAC1 et MBCAC	132
I.1 γ_s^{DL} (en dB) en fonction de FER_s^{DL} et R_s^{DL}	145
I.2 γ_s^{UL} (en dB) en fonction de FER_s^{UL} et R_s^{UL}	146

LISTE DES FIGURES

1.1	Schéma simplifié d'un réseau de téléphonie mobile	3
1.2	Réseaux cœurs envisagés et leur interconnexion avec les réseaux d'accès radio CDMA2000 et UMTS	6
1.3	Architecture du réseau CDMA2000 à commutation de paquets sous IP mo- bile	7
1.4	Les différentes techniques d'accès multiple	11
1.5	Différents types d'accès multiple à partir de la technique OFDM	11
1.6	Etalement de spectre par séquence directe	13
1.7	Principe du <i>soft handover</i> dans un système CDMA	18
1.8	Architecture d'un nœud <i>Integrated Services</i>	25
1.9	Protocole de réservation RSVP	26
1.10	Ordonnanceur à isolation de flux	29
1.11	Token Bucket	33
1.12	Caractérisation d'un trafic	34
1.13	Courbes d'arrivée et de service	36
2.1	Configuration des cellules sans <i>soft handover</i>	66
2.2	Configuration des cellules avec <i>soft handover</i>	67
2.3	Influence du rayon	68
2.4	Géométrie des cellules	69
2.5	Influence du type de trafic	70
3.1	Architecture du réseau modélisé	74
3.2	Modélisation des files d'attente	76
3.3	Délai de transmission radio	78
3.4	Composantes du délai d'un paquet transmis sur le réseau cœur	79
3.5	Evolution du débit alloué à la session s lors de sa période d'activité $[0,t)$.	83
3.6	Mécanisme de mesure	95

3.7	Pseudo-code de l'algorithme de mesure	96
3.8	Algorithme MBCAC sens montant	102
3.9	Algorithme MBCAC sens descendant	103
4.1	Topologie de l'environnement simulé	105
4.2	Succession temporelle des profils de trafic	110
4.3	Succession temporelle des sessions pour un profil de trafic considéré	111
4.4	Quantité de trafic pour le scénario (A)	115
4.5	Profil de trafic pour le scénario (A)	115
4.6	Quantité de trafic pour le scénario (B)	116
4.7	Profil de trafic pour le scénario (B)	117
4.8	Taux de blocage globaux obtenus pour 5 séquences de simulations selon le scénario A avec la procédure CAC1 et selon le scénario B avec la procédure MBCAC	118
4.9	NdS globaux en fonction des paramètres de mesure pour les scénarios A et B	121
4.10	Taux d'utilisation de la capacité disponible à la BS en fonction des pa- ramètres de mesure pour les scénarios A et B	122
4.11	Distributions cumulatives de la répartition du délai bout en bout pour les applications <i>Voice Gold</i> et <i>DB Gold</i> en fonction des paramètres de mesure	124
4.12	Mesures du trafic DB dans le sens descendant en fonction des paramètres de mesure	126
4.13	NdS globaux, suivant le profil d'utilisateur et le type d'application, pour le scénario A sous les procédures CAC1 et MBCAC	129
4.14	NdS globaux, suivant le profil d'utilisateur et le type d'application, pour le scénario B sous les procédures CAC1 et MBCAC	130
4.15	Taux d'utilisation de la capacité disponible à la BS pour les scénarios A et B sous les procédures CAC1 et MBCAC	132
4.16	Distributions cumulatives de la répartition du délai bout en bout pour les applications <i>Voice Gold</i> et <i>DB Gold</i> sous les procédures CAC1 et MBCAC .	133

4.17 Taux de blocage en fonction du rayon de la cellule pour le scénario B sous la procédure MBCAC	134
II.1 Schéma des cellules pour le calcul des coefficients d'atténuation	150
III.1 Protocole de répétition sélective	153
IV.1 Algorithme CAC0 sens montant	154
IV.2 Algorithme CAC0 sens descendant	155
IV.3 Algorithme CAC1 sens montant	156
IV.4 Algorithme CAC1 sens descendant	157

LISTE DES NOTATIONS ET DES SYMBOLES

- 3G : Third Generation Cellular Mobile Systems
- AAA : Authentification, Autorization, and Accounting
- AMPS : Advanced Mobile Phone Services
- ANSI : American National Standards Institute
- ARQ : Automatic Request for Retransmission
- BBRR : Bit by Bit Round Robin
- BE : Best Effort
- BS : Station de Base
- BSC : Contrôleur de Station de Base
- CAC : Call Admission Control
- CDMA : Code Division Multiple Access
- CDG : CDMA Development Group
- CN : Core Network
- DL : Downlink
- DS : Direct Spread
- FER : Frame Error Rate
- FIFO : First In, First Out
- FA : Foreign Agent
- FCH : Fundamental CHannel
- FDD : Frequency Division Duplex
- FDMA : Frequency Division Multiple Access
- FQ : Fair Queueing
- GPS : Generalized Processor Sharing
- GSM : Global System for Mobile Communication
- HA : Home Agent

- IETF : Internet Engineering Task Force
- IMT : International Mobile Telecommunications
- IntServ : Integrated Services
- IP : Internet Protocol
- ISP : Internet Service Provider
- MBCAC : Measurement-Based Admission Control
- MC : Multi-Carrier
- NdS : Niveau de service
- OFDM : Orthogonal Frequency Division Multiplexing
- PCF : Packet Control Function
- PDSN : Packet Data Serving Node
- PGPS : Packet by Packet Generalized Processor Sharing
- PN : Pseudorandom Noise
- PPP : Point to Point Protocol
- QoS : Qualité de Service
- RAB : Radio Access Bearer
- RN : Radio Network
- RNIS : Réseau Numérique Intégré de Services
- RPPS : Rate Proportional Processor Sharing
- RRC : Radio Resources Control
- RSPEC : Ressources SPECification
- RSVP : Resource reSerVation setup Protocol
- SCH : Supplemental CHannel
- SIR : Signal to Interference Ratio
- SMS : Short Messages Services
- SRP : Protocole de Répétition Sélective
- TDD : Time Division Duplex
- TDMA : Time Division Multiple Access

- TIA : Telecommunication Industry Association
- TSPEC : Traffic SPECification
- UIT : Union Internationale des Télécommunications
- UL : UpLink
- UMTS : Universal Mobile Telecommunication System
- VoIP : Voix sur IP
- WFQ : Weighted Fair Queuing

LISTE DES ANNEXES

Annexe I:	$\frac{E_b}{N_t}$ requis en fonction du RAB alloué	145
I.1	Abaques pour le sens descendant	145
I.2	Abaques pour le sens montant	146
Annexe II:	Calcul des coefficients d'atténuation pour les formules de	
	capacité radio dans le sens descendant	147
II.1	Modèle de propagation	147
II.2	Calcul des coefficients $E[a_{ms,0}]$ et $E[\frac{a_{ms,0}}{a_{ms,i}}]$	147
II.3	Calcul des coefficients $a_{m*,0}$ et $a_{m*,i}$	149
Annexe III:	Les mécanismes de retransmission de l'ARQ	152
III.1	<i>Stop and wait</i>	152
III.2	<i>Go-back-N</i>	152
III.3	Répétition sélective	153
Annexe IV:	Les algorithmes d'admission CAC0 et CAC1	154

INTRODUCTION

0.1 Motivations

L'essence de la technologie de troisième génération est une combinaison entre la transmission des données par paquets, un réseau fondé entièrement sur le protocole IP (*Internet Protocol*) et une nouvelle interface radio basée sur la technologie CDMA2000 (pour l'Amérique du Nord).

Le défi de cette nouvelle ère de communication mobile est de pouvoir accéder à des services multimédias en tout lieu et en tout temps. Les réseaux cellulaires de troisième génération sont ainsi le point de convergence entre trois secteurs principaux des technologies de l'information: les applications multimédias, l'Internet et l'informatique mobile.

L'Internet actuel s'appuie sur un protocole non orienté connexion et ne fournit qu'un service "au mieux" (*best-effort*), c'est-à-dire qu'aucune garantie d'aucune sorte n'est faite sur le service offert, autre que l'engagement du réseau à faire "du mieux qu'il peut" pour amener les paquets à destination. Ce type de service était adapté pour les applications du début de l'Internet, telles que la navigation ou FTP, qui n'exigent qu'un transport fiable des données sans contraintes temporelles et peuvent donc accepter de grandes variations de délai et compenser les pertes éventuelles par des retransmissions.

Cependant, le développement de l'Internet a suscité la création de nouveaux types d'applications, multimédias, qui ont des contraintes de type temps réel et ont besoin de garanties de délai et de débit, nécessitant une transmission des paquets suivant un mode orienté connexion. En particulier, ces applications ont besoin de hauts débits de transmission et exigent des contraintes de délai strictes.

Enfin, cette nouvelle demande de service représente également un défi pour l'interface radio qui doit pouvoir maintenir des hauts débits et une garantie de qualité de service sur les liens radios qui sont peu fiables. La troisième génération de mobiles utilise ainsi la technique de multiplexage CDMA qui permet d'augmenter la capacité des cellules radios et de proposer des débits diversifiés.

Les caractéristiques spécifiques à chaque domaine rendent donc difficile la mise en

place de réseaux de communication multimédias mobiles. De nouveaux mécanismes doivent être employés afin de différencier les services et fournir la qualité de service exigée. Il s'agit notamment d'utiliser la réservation des ressources, des politiques d'ordonnancement des paquets adéquats, et une politique de contrôle d'admission stricte doit être mise en place sur l'ensemble du réseau.

Notre étude se donne ainsi pour objectif de développer une politique de contrôle d'admission pour un réseau de troisième génération à contraintes de service différenciées. Le contrôle d'admission se fait dans le contexte d'un réseau à commutation de paquets, et les décisions d'admission ou de rejet se font indépendamment pour chaque flot en entrée du réseau (réseau cœur ou réseau d'accès radio). Nous proposons une politique d'admission adaptative qui réévalue les ressources allouées aux flots déjà admis dans le réseau, puis nous y ajoutons un algorithme de mesure afin de bénéficier d'un gain en termes d'utilisation de la bande passante. L'impact des algorithmes sur la performance du système est ensuite étudié.

0.2 Organisation du mémoire

Le chapitre 1 présente le contexte des réseaux de troisième génération. La technique d'accès radio CDMA y est décrite en vue d'étudier les mécanismes influençant la capacité d'une cellule radio. La discipline d'ordonnancement WFQ susceptible d'être déployée dans les nœuds du réseau cœur est également étudiée afin d'obtenir une relation entre les ressources disponibles et la garantie de QoS offerte dans le réseau.

Le chapitre 2 propose une nouvelle estimation de la capacité radio CDMA2000 pour le sens montant et le sens descendant dans un contexte multiservices. Les formules obtenues sont ensuite validées numériquement.

Le chapitre 3 définit la politique de contrôle d'admission bout en bout développée à partir des formules de capacité radio élaborées au chapitre précédent d'une part, et des propriétés de la discipline WFQ sur les nœuds du réseau cœur d'autre part. Il présente d'abord un algorithme basé sur des descripteurs de trafic à priori puis celui à base de mesures.

Le chapitre 4 présente les différentes expérimentations menées afin de déterminer les caractéristiques de l'algorithme de contrôle d'admission à base de mesures, puis de le comparer à un algorithme basé sur des descripteurs de trafic à priori. Des simulations montrent également l'utilisation possible de l'algorithme d'admission pour effectuer du dimensionnement radio.

Enfin une conclusion est apportée quant à l'ensemble des résultats obtenus, et les améliorations envisageables de l'algorithme d'admission sont évoquées.

0.3 Contributions

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

- une estimation des ressources radios nécessaires à la satisfaction des exigences de QoS des sessions,
- une procédure de contrôle d'admission bout en bout prenant en compte à la fois l'admission dans le réseau cœur et le réseau radio,
- une politique de contrôle d'admission à base de mesures adaptative, susceptible de prévenir l'apparition du phénomène de congestion dans une cellule radio et de privilégier les sessions de priorités élevées par rapport aux sessions de priorités faibles lors de l'admission,
- un simulateur prototype pour l'évaluation des performances relatives à la conception et au dimensionnement des futurs réseaux de troisième génération.

CHAPITRE 1

ARCHITECTURE ET SERVICES DES RÉSEAUX 3G

1.1 Présentation générale des réseaux de troisième génération

Les systèmes cellulaires de téléphonie mobile actuels et leurs évolutions sont présentés dans cette section afin de pouvoir se repérer parmi les différents systèmes existants. Les spécificités des systèmes cellulaires de troisième génération sont ensuite analysées.

1.1.1 Historique des systèmes cellulaires

L'évolution des systèmes cellulaires est classée en trois grands groupes: la première génération qui a démarré à la fin des années 1970 et qui existe encore de nos jours; la deuxième génération qui a fait ses débuts dans les années 1990 et qui continue à se développer; et enfin, les systèmes de troisième génération dont les premiers déploiements commerciaux sont prévus en ce début de siècle. D'un point de vue commercial, on assiste à la fin de la première génération de systèmes cellulaires aussi appelés systèmes "analogiques" et la deuxième génération composée des systèmes dits "numériques" est en plein essor. Mais la deuxième génération est victime de son succès, car le nombre d'abonnés ne cesse de croître et le résultat est une saturation du spectre de fréquence.

Les systèmes de première génération ont été les premiers à exploiter le concept de la téléphonie cellulaire développé par les laboratoires Bell en collaboration avec quelques industriels américains dans les années 1960-1970. Ces systèmes ont la particularité d'utiliser la modulation en fréquence (qui est de type analogique) et uniquement la méthode d'accès multiple par répartition de fréquences (FDMA, voir la section 1.3.1.1). C'est le système américain AMPS (*Advanced Mobile Phone Services*), introduit en 1983, qui est considéré comme l'exemple le plus représentatif de cette première génération. En termes de services, ces systèmes proposaient de la voix qui était transmise sous forme analogique, avec une efficacité spectrale assez médiocre.

Les systèmes de deuxième génération ont profité des développements technologiques

réalisés dans le domaine des composants radiofréquence et des dispositifs de traitement numérique du signal. A la différence des systèmes de première génération, les systèmes de deuxième génération utilisent des techniques numériques de codage de la parole et la modulation est elle aussi numérique. De plus, ces systèmes ont été conçus pour proposer non seulement de la voix mais aussi d'autres services tels que l'envoi de messages (SMS pour *Short Messages Services*), le fax et l'accès à des réseaux haut débit tels que l'Internet et le RNIS (Réseau numérique Intégré de Services). Le tableau 1.1 décrit les caractéristiques techniques radio des systèmes cellulaires actuels de deuxième génération en Europe et aux Etats-Unis.

TAB. 1.1 – *Caratéristiques d'accès radio des principaux systèmes cellulaires de deuxième génération. VD: voie descendante, VM: voie montante, QPSK: quadrature phase shift keying, GMSK: gaussian minimum shift keying*

Standard	IS-95 (CDMAOne)	GSM
Pays d'origine	Etats-Unis	Europe
Lancement commercial	1995	1992
Technique d'accès multiple	FDMA/CDMA	FDMA/TDMA
Spectre des fréquences (MHz)	869-894(VD) 824-849(VM) 1850-1910(VD) 1930-1990(VM)	925-960(VD) 880-915(VM) 1805-1880(VD) 1720-1785(VM) 1850-1910(VD) 1930-1990(VM)
Mode de duplexage	FDD	FDD
Séparation entre porteuses	1250kHz	200kHz
Type de modulation	QPSK	GMSK

Les références techniques données dans cette section sont précisées dans l'ouvrage de Sanchez^[1].

1.1.2 Systèmes cellulaires de troisième génération

La troisième génération un ensemble de technologies développées dans le but de faire évoluer les systèmes cellulaires actuels de deuxième génération, tant en termes de capacité et de couverture, que de variété et de qualité de services. Les mobiles de troisième génération ne transporteront pas seulement de la voix, ils seront de véritables terminaux multimédias capables d'offrir simultanément des services de transmission de données, d'audio et de vidéo en tout lieu et à tout moment. C'est en termes de services que

la troisième génération se démarque surtout de la deuxième génération. D'une part la 3G autorise des débits de transmission plus élevés, d'autre part il est possible dans la 3G d'établir et de maintenir plusieurs connexions simultanément avec la possibilité de négocier le type de trafic et ses caractéristiques de manière dynamique, ce qui est apprécié pour offrir des services multimédias.

Afin de comprendre l'évolution des réseaux cellulaires vers la troisième génération, nous devons décrire brièvement les éléments qui les composent. On observe, sur la figure 1.1, qu'un réseau de téléphonie mobile est constitué des terminaux mobiles, d'un réseau d'accès radio RN (*radio network*) et d'un réseau de cœur CN (*core network*). Le réseau cœur est celui qui est en charge de l'acheminement des communications vers les réseaux fixes tels que le réseau public de téléphonie fixe, le réseau Internet, etc. Enfin, le rôle du réseau d'accès radio est d'acheminer l'information depuis le terminal mobile jusqu'au réseau cœur.

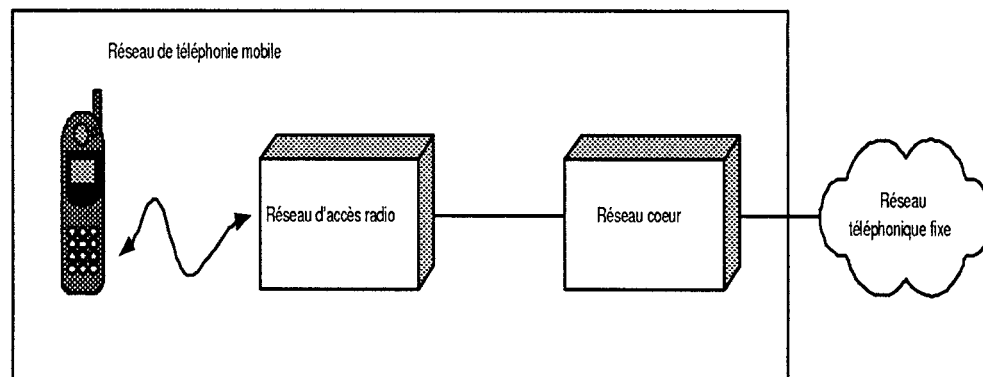


FIG. 1.1 – Schéma simplifié d'un réseau de téléphonie mobile

1.1.2.1 Les réseaux d'accès radio de troisième génération

L'UIT (Union Internationale des Télécommunications), organisme chargé de réglementer les différents standards de télécommunications au niveau mondial, a voulu mettre sur pied un standard unique et international de système cellulaire de troisième génération. Cette norme "unique" fut désignée par le sigle IMT-2000 (pour *International Mobile Telecommunications 2000*). Le problème est qu'en 1998, quinze technologies d'accès radio ont

été proposées à l'UIT par les organismes de standardisation des différents pays membres. Finalement, l'IMT-2000 ne couvrira pas un seul système cellulaire mais un ensemble de systèmes de troisième génération. Nous nous intéressons dans ce mémoire uniquement au système CDMA2000 (Etats-Unis), tout en le situant par rapport au standard UMTS (Europe). Il existe six technologies d'accès radio terrestre ayant le label IMT-2000, dont les technologies UTRA/FDD et UTRA/TDD que regroupe l'UMTS, et la technologie de CDMA2000.

La technologie CDMA2000 (ou IS-2000) a été soutenue par le consortium CDG (*CDMA Development Group*) dans le cadre de l'évolution de l'IS-95 vers la troisième génération. Elle offre la possibilité d'utiliser une ou plusieurs porteuses à l'aide de la méthode d'accès MC-CDMA (voir section 1.3.1.1). Ainsi, l'évolution du CDMAOne (ou IS-95) vers la troisième génération se fera en deux phases: dans la première, le CDMA2000, appelé "CDMA2000 phase 1" ou "CDMA2000 1X", sera capable d'offrir, à l'aide d'une seule porteuse avec une bande de 1.25 MHz, des débits allant jusqu'à 144 kbps. Il sera directement compatible avec le CDMAOne. Dans une deuxième phase, le "CDMA2000 phase2" ou "CDMA2000 3X", permettra, à l'aide de trois porteuses avec une bande de 3×1.25 MHz, d'atteindre un débit de 2 Mbps. La principale similitude technique entre le CDMA2000 et les technologies de l'UMTS est l'utilisation du CDMA comme mode d'accès (voir section 1.3.1.1). On remarque cependant que le CDMA2000 fait appel à la modulation multiporteuse dans la voie descendante (jusqu'à douze porteuses simultanément), ce qui s'avère particulièrement utile pour garder de la compatibilité avec le CDMAOne. Un travail d'harmonisation entre le CDMA2000 et l'UMTS est en cours et à l'avenir, ces deux systèmes seront capables de réaliser une procédure d'itinérance (*roaming*) entre eux (CDMA2000 vers UMTS et *vice versa*). Les paramètres techniques sur lesquels reposent les technologies d'accès radio de troisième génération de l'UMTS et de CDMA2000 sont présentés dans le tableau 1.2.

Agréés par l'UIT, les réseaux d'accès 3G doivent satisfaire les caractéristiques techniques suivantes:

- services à haut débit. Minimum 144 kbps dans tout type d'environnement et 2 Mbps dans les environnements intérieurs et avec une mobilité restreinte;

TAB. 1.2 – *Caratéristiques des technologies d'accès radio de CDMA2000 et de l'UMTS (VD:voie descendante, VM: voie montante).*

Standard	CDMA2000	UMTS	
Pays d'origine	Etats-Unis	Europe	
Technologie d'accès radio	CDMA2000	UTRA/FDD ou WCDMA	UTRA/TDD
Technique d'accès multiple	CDMA(VM) MC-CDMA(VD)	CDMA	TDMA et CDMA
Mode de duplexage	FDD	FDD	TDD
Séparation entre porteuses	$N \times 1.25$ MHz où $N=1,3,6,9,12$	5 MHz	5 MHz
Synchronisation entre stations de base	Synchrone	Asynchrone (Synchrone optionnelle)	Synchrone
Durée d'une trame (ms)	5,10,20,40,80	10	10
Type de modulation des données	BPSK(VM) QPSK(VD)	BPSK(VM) QPSK(VD)	QPSK
Site avec les spécifications	www.3gpp2.org/Public_html/specs/index.cfm www.tiaonline.org/standards	www.3gpp.org/specs/specs.htm	www.3gpp.org/specs/specs.htm

- transmission de données de manière symétrique et asymétrique. Dans une transmission symétrique, le débit de la voie montante (du terminal mobile vers le réseau fixe) est égal à celui de la voie descendante (du réseau fixe vers le terminal mobile). Au contraire, dans une communication asymétrique, le débit dans les deux voies est différent;
- services à commutation de paquets pour la transmission de données, et commutation de circuit pour la transmission de la voix;
- qualité de la parole comparable à celle des réseaux câblés;
- capacité et efficacité spectrale supérieures à celles des systèmes cellulaires actuels de deuxième génération;
- possibilité d'offrir des services multimédias lors d'une même connexion et avec des qualités de services différentes (débit, taux d'erreurs, délai de transfert...) pour les différents types de médias (voix, audio, données...);
- compatibilité avec les réseaux d'accès radio de deuxième génération;
- itinérance (*roaming*) entre les différents systèmes de troisième génération.

1.1.2.2 Les réseaux cœurs des systèmes de troisième génération

La figure 1.2 représente les trois réseaux de cœur possibles actuellement envisagés pour les systèmes de troisième génération IMT-2000. On voit que les réseaux de cœur utilisés actuellement dans la deuxième génération seront également employés dans la troisième génération sous une forme évoluée: ainsi ces réseaux cœurs s'appuient sur les protocoles GSM/MAP (en Europe) et l'ANSI-41 (aux Etats-Unis). Cependant il est question d'adopter plus tard un réseau cœur entièrement IP (*Internet Protocol*), ce qui facilitera le *roaming* entre les différents systèmes de troisième génération.

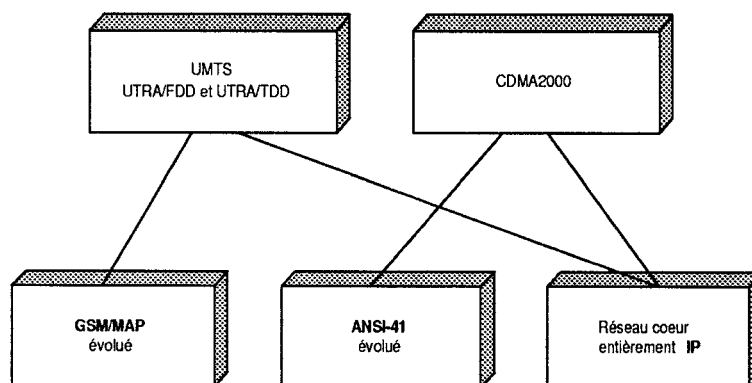


FIG. 1.2 – Réseaux cœurs envisagés et leur interconnexion avec les réseaux d'accès radio CDMA2000 et UMTS

1.2 Le réseau cœur CDMA2000

1.2.1 Architecture

Le réseau cœur interconnecte les réseaux périphériques. Il peut être le réseau d'un ISP (*Internet Service Provider*) ou le réseau interne d'une entreprise. Il appartient à un domaine d'administration dans lequel est définie une politique de Qualité de Service (QoS) qui sera mise en œuvre dans les routeurs. Ce réseau est constitué de routeurs de cœur dont la fonction est de router les paquets en fonction des informations de Qualité de Service qu'ils contiennent.

Contrairement aux systèmes précédents, le réseau cœur des systèmes de troisième génération verra son architecture complètement basée sur la commutation de paquets suivant

le protocole IP, avec le protocole VoIP pour la voix. L'architecture de ce futur réseau cœur à commutation de paquets est décrite par le standard IS-835-B^[2] de l'organisme américain TIA (*Telecommunications Industry Association*). Cette norme repose sur le RFC 2002^[3] de l'IETF (pour *Internet Engineering Task Force*). La figure 1.3 représente l'architecture du réseau CDMA2000 proposé dans cette norme.

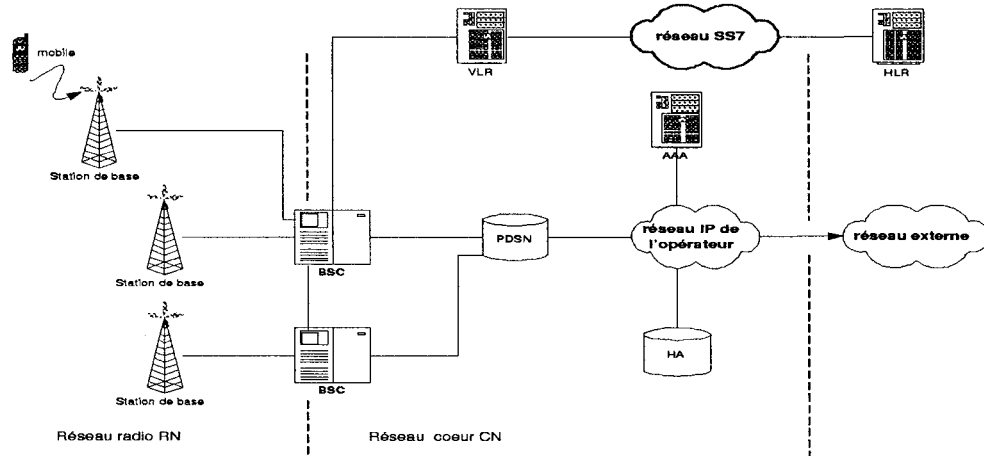


FIG. 1.3 – Architecture du réseau CDMA2000 à commutation de paquets sous IP mobile

Les éléments importants dans l'architecture du réseau sont les suivants:

- BSC (*Base Station Controller*). Le BSC est un élément du réseau contenant deux composants logiques: la fonction de contrôle des paquets (PCF pour *Packet Control Function*) et le contrôle des ressources radio (RRC pour *Radio Resources Control*). Le PCF est en charge de l'établissement, du maintien et de la fin d'une connexion avec le PDSN. Il communique également avec le RRC afin de gérer les ressources radio pour la transmission sur le lien radio. Lors d'un *soft handover* avec un BSC relié à une autre station de base, le PCF transfère ses informations au nouveau PCF pour l'établissement d'une nouvelle transmission des paquets vers le PDSN. Le RRC quant à lui gère les ressources radio pour chaque usager, en maintenant le transfert des paquets entre le terminal mobile et le PCF.
- PDSN (*Packet Data Serving Node*). Le PDSN intègre beaucoup de fonctionnalités. Il est en charge du routage des paquets vers le réseau IP ou directement vers le HA (*Home Agent*). Il affecte également les adresses IP dynamiques et maintient les

liens PPP (*Point to Point Protocol*) vers les terminaux mobiles. Il transfère vers le serveur AAA les données d'identification propres de l'abonné. En retour le PDSN reçoit du AAA les paramètres du profil de l'utilisateur (services différenciés et sécurité). Le PDSN peut aussi jouer le rôle de FA (*Foreign Agent*) pour la mobilité IP.

- AAA (*Authentication, Authorization and Accounting*). Le serveur AAA contient les informations relatives aux abonnés nécessaires à l'identification et à la définition des paramètres d'une session utilisateur. Il permet aussi de fournir les indices de tarification collectés par le nœud PDSN pour l'utilisateur considéré.

1.2.2 Mobilité IP

Le concept de mobilité IP a été développé dans le RFC 2002^[3] et constitue le fondement de l'architecture du réseau cœur décrite dans la section précédente. L'adresse IP d'un nœud identifie le point d'accès au réseau et est utilisée par les stratégies de routage pour accéder au nœud. Or, un nœud mobile, par définition, change de point d'accès et il ne peut donc pas avoir une adresse fixe. Ainsi on donne au nœud un identificateur valable dans tout le réseau et on établit une correspondance entre l'adresse et l'identificateur. D'autre part, il faut pouvoir maintenir les connexions de l'hôte mobile lors de son déplacement et cela de façon complètement transparente pour les applications et l'utilisateur. Un nœud mobile peut être adressé de deux façons:

- en utilisant son adresse maison (*home address*), adresse IP permanente indépendante de la localisation courante du nœud mobile; l'adresse maison est relative au sous-réseau IP maison (d'appartenance) du nœud mobile;
- en utilisant son adresse mobile (*care-of address*), adresse IP temporaire dépendante de la localisation courante du nœud mobile; l'adresse mobile est relative au sous-réseau IP de raccordement du nœud mobile; cette adresse mobile peut être obtenue par des mécanismes d'autoconfiguration d'adresse.

On s'intéresse au cas où le nœud mobile n'est pas raccordé à son réseau maison. Un émetteur peut envoyer des paquets au nœud mobile soit directement en utilisant l'adresse mobile, s'il la connaît, soit indirectement en utilisant l'adresse maison. En effet, le nœud mobile est toujours accessible par son adresse maison.

Quand un nœud mobile quitte le réseau maison pour être rattaché à un réseau étranger (*foreign network*), il obtient d'abord une adresse mobile et prévient ensuite un routeur du réseau maison du changement (déclaration de position). Ce routeur s'appelle agent maison HA (*home agent*) du nœud mobile. L'agent maison doit maintenir pour chaque nœud mobile enregistré une entrée dans une table de correspondance (*home address* $\leftrightarrow$ *care-of address*). Cette entrée contient l'adresse mobile (*care-of address*), un numéro d'identification et une durée de vie de la correspondance.

L'émetteur connaît l'adresse maison du nœud mobile. Il émet son paquet à cette adresse. L'agent maison intercepte ce paquet, consulte sa table, détermine la localisation du nœud mobile, et reroute alors le paquet vers le nœud mobile. Le nœud mobile s'aperçoit alors que le paquet n'a pas été envoyé directement par l'émetteur et lui envoie un paquet de mise à jour de sa table de correspondance. L'émetteur connaît ainsi l'adresse mobile et va l'utiliser pour ses prochains paquets.

1.3 Le réseau d'accès radio

Les techniques principales utilisées dans le réseau d'accès radio sont décrites dans cette section, comme cela est détaillé dans l'ouvrage de Sanchez^[1] et de Holma et Toskala^[4].

1.3.1 La méthode d'accès radio CDMA

1.3.1.1 Rappel sur les méthodes d'accès multiple

Le concept d'accès multiple est le fondement de tout système radio mobile. Il s'agit de partager un ensemble limité de canaux de communication, de telle sorte que plusieurs utilisateurs puissent y avoir accès pour communiquer simultanément. Un canal est donc une portion de cet ensemble qui est alloué temporairement à un utilisateur pour sa communication. Il est donc important de rentabiliser au maximum cette ressource, et dans cet objectif de nouvelles techniques d'accès multiple sont mises au point afin d'accroître la capacité et la couverture des systèmes cellulaires tout en préservant la qualité de service.

D'abord on rappelle que toute technique d'accès multiple est assortie d'un type particulier de duplexage permettant l'échange d'information entre le mobile et la sta-

tion de base. Un système de communication bidirectionnel grâce auquel on peut transmettre et recevoir des informations simultanément est appelé système *full-duplex* (le plus fréquent). Parmi les méthodes de type *full-duplex*, on distingue les techniques de duplexage en fréquence (FDD pour *Frequency Division Duplex*) et celles qui effectuent ce duplexage dans le temps (TDD pour *Time Division Duplex*). Le mode FDD utilise deux bandes de fréquences indépendantes (l'une pour transmettre et l'autre pour recevoir simultanément). Dans le mode TDD, on transmet et on reçoit sur la même bande de fréquence mais à des instants différents.

La première technique d'accès multiple est l'accès multiple par répartition de fréquences (FDMA pour *Frequency Division Multiple Access*). En FDMA, on alloue à chaque utilisateur une bande fréquence unique, comme le montre la figure 1.4.a (si le mode de duplexage est FDD alors on alloue en réalité deux bandes de fréquence: l'une pour transmettre et l'autre pour recevoir). Lorsqu'un usager souhaite établir une communication, son signal est modulé par une fréquence porteuse qui lui est propre. Ainsi la ressource radio élémentaire dans un système FDMA est une largeur spectrale finie.

La deuxième technique d'accès multiple utilisée est l'accès multiple par répartition dans le temps (TDMA pour *Time Division Multiple Access*). En TDMA, une même bande de fréquence est partagée par un certain nombre d'utilisateurs qui se voient attribuer un intervalle (ou *slot*) de temps unique, comme le montre la figure 1.4.b. Le TDMA est souvent utilisé combiné avec le FDMA car il permet d'augmenter la capacité d'un système FDMA en appliquant du TDMA sur chacune des fréquences porteuses.

La dernière technique d'accès multiple est l'accès multiple par répartition de codes, appelée CDMA (pour *Code Division Multiple Access*). En CDMA, les utilisateurs peuvent communiquer simultanément dans une même bande de fréquence. La distinction entre les différents utilisateurs s'effectue alors grâce à un code qui leur est attribué et connu exclusivement par l'émetteur et le récepteur. Le principe du CDMA est représenté dans la figure 1.4.c. Si l'on considère les systèmes FDMA et TDMA, on s'aperçoit que leur capacité est limitée par la bande de fréquence allouée à chaque utilisateur. Dans la section suivante on verra que la capacité d'un système fondé sur le CDMA est limitée notamment par le niveau d'interférences dans le réseau.

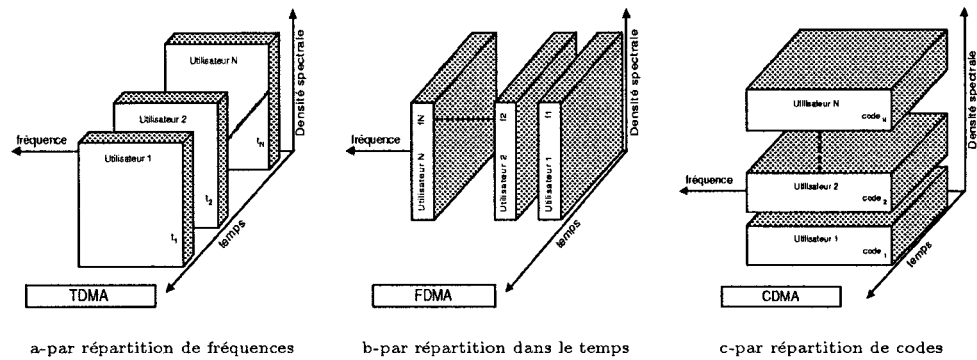


FIG. 1.4 – Les différentes techniques d'accès multiple

Il existe également le multiplexage par répartition sur des fréquences orthogonales (OFDM pour *Orthogonal Frequency Division Multiplexing*) qui est un cas particulier de la modulation multiporteuse (MC pour *Multi-Carrier*). Un bloc de symboles d'information est décomposé en sous-blocs. Chaque sous-bloc est ensuite transmis sur différentes sous-porteuses en parallèle à un débit inférieur. Les signaux OFDM doivent être orthogonaux, de telle sorte que les sous-porteuses puissent se recouvrir dans un même spectre. Dans le contexte de la téléphonie cellulaire, l'OFDM peut être utilisé en combinaison avec d'autres formes d'accès multiple comme le FDMA, le TDMA et le CDMA pour donner lieu, respectivement aux systèmes MC-FDMA, MC-TDMA, et MC-CDMA. Ces trois formes combinées sont représentées dans la figure 1.5. La technique MC-CDMA a été adoptée dans le système de troisième génération CDMA2000 (pour le sens descendant) que nous allons considérer dans ce mémoire.

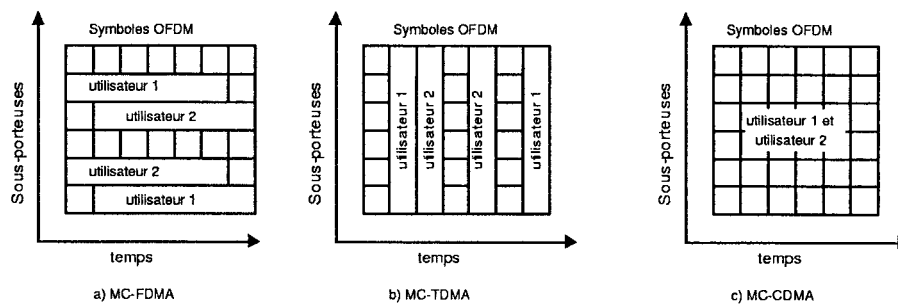


FIG. 1.5 – Différents types d'accès multiple à partir de la technique OFDM

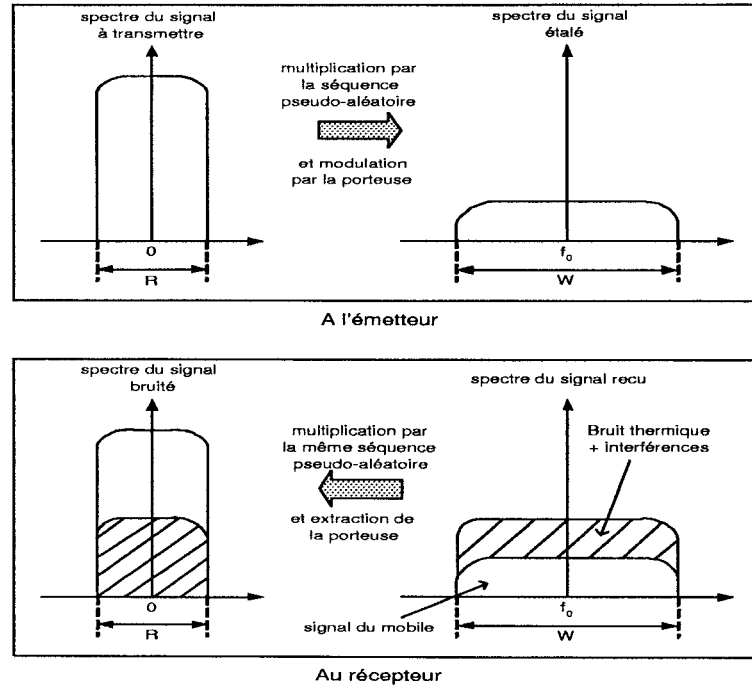
1.3.1.2 Étalement de spectre et DS-CDMA: Principe

Le CDMA est aussi connu sous le nom d'accès multiple par étalement de spectre. En effet, l'étalement de spectre est la technique sur laquelle repose le CDMA et qui permet à plusieurs utilisateurs d'être présents simultanément sur une même bande de fréquence. Cette technique permet de transmettre un signal d'information sur une largeur de bande plusieurs fois supérieure à la largeur de bande nécessaire pour transmettre le signal. Dans un système à étalement de spectre, le signal transmis est "étalé" à partir d'un code indépendant du message d'information. Après s'être synchronisé avec l'émetteur, le récepteur doit utiliser ce même code pour "désétaler" le signal et pouvoir par la suite récupérer le message d'information.

Le DS-CDMA ou CDMA à séquences directes est la technique d'étalement la plus répandue, elle est utilisée dans le système de deuxième génération CDMAOne actuel et sera également utilisée en CDMA2000. La méthode consiste à multiplier le signal binaire d'information par une séquence binaire pseudo-aléatoire (code PN pour *Pseudorandom Noise*) dont la fréquence des symboles (appelés *chips*) est beaucoup plus élevée que la fréquence des symboles du signal. Cela a donc pour effet d'étaler la largeur de bande du signal. De même la puissance du signal se retrouve répartie sur toute la nouvelle largeur de bande. Le signal se retrouve alors noyé dans le "bruit" (ce bruit est constitué de toutes les sources d'interférences et du bruit thermique). A l'arrivée, le signal étalé est remultiplié par le même code PN. Les caractéristiques particulières des codes PN (notamment l'orthogonalité entre eux) font que l'on peut ainsi retrouver le signal de départ. En effet, le signal arrive au récepteur noyé dans le bruit et la multiplication par la séquence pseudo-aléatoire permet d'extraire le signal car seul le signal qui avait été multiplié au départ par cette séquence verra sa largeur de bande réduite, tandis que le bruit restera étalé sur la largeur de bande totale (figure 1.6).

Ainsi on constate que le signal "étalé" résiste fort bien aux interférences. C'est grâce à cette robustesse vis-à-vis des interférences que l'accès multiple est possible.

Cependant, même si les usagers sont séparés entre eux par leur code PN respectif, une autre condition est nécessaire pour assurer que le message transmis puisse être décodé

FIG. 1.6 – *Étalement de spectre par séquence directe*

avec un minimum d'erreur: le rapport $\frac{E_b}{N_t}$ du signal à la réception doit être suffisamment élevé. E_b représente l'énergie par bit du signal reçu (en Joule) et N_t la densité spectrale du bruit (interférences+bruit thermique) (en Watt/Hz). Ce rapport est directement lié au rapport signal à bruit SIR (pour *Signal to Interference Ratio*) qui est le rapport de la puissance du signal reçu sur la puissance du bruit (interférences+bruit thermique):

$$\frac{E_b}{N_t} = \frac{W}{R} \text{SIR}.$$

On met ainsi en évidence un paramètre clé de tout système d'accès radio à étalement de spectre qui est le gain de traitement (*processing gain*). Ce dernier est défini comme le rapport de la largeur de bande occupée par le signal d'information après (largeur de bande W) et avant (largeur de bande R) étalement. Ainsi, on voit que plus le rapport $\frac{W}{R}$ est grand (donc plus le signal est étalé), plus on peut se permettre un rapport signal à bruit petit tout en assurant un $\frac{E_b}{N_t}$ suffisamment élevé, d'où l'importance de ce gain de traitement.

Il s'agit maintenant de préciser d'où proviennent les interférences. On distingue les interférences intra-cellulaires et les interférence inter-cellulaires.

Interférences intra-cellulaires. Elles représentent l'interférence mutuelle entre les utilisateurs de la cellule. Des codes orthogonaux sont utilisés à la fois dans la voie descendante et dans la voie montante, et si cette orthogonalité était préservée alors les signaux des différents usagers de la cellule seraient décorrélés entre eux et il n'y aurait pas d'interférences intra-cellulaires. Dans la voie descendante, en absence de trajets multiples, les signaux gardent leur orthogonalité car ils sont transmis alignés dans le temps par la station de base: les codes sont synchronisés en temps (les stations de base respectent en effet une référence de temps unique pour transmettre, cette référence de temps pouvant être fournie par un système de navigation par satellite). Dans la réalité on a toujours des trajets multiples, c'est-à-dire que plusieurs copies du signal transmis arrivent au récepteur à des instants différents. Du fait de ces trajets multiples, les codes dans la voie descendante ne restent pas parfaitement orthogonaux, et on introduit donc dans ce sens un facteur d'orthogonalité w , $w=0$ correspondant à une orthogonalité parfaite et pas d'interférences intra-cellulaires, $w=1$ correspondant au fait que tous les signaux de la cellule interfèrent pleinement entre eux. A la différence de la voie descendante, dans la voie montante, toujours en absence de trajets multiples, les signaux des différents utilisateurs de la cellule ne restent pas orthogonaux car les utilisateurs de la cellule transmettent de façon indépendante et non synchronisée. Ainsi, dans le sens montant, les signaux interfèrent pleinement avec ou sans trajets multiples.

Interférences inter-cellulaires. Dans le sens montant, l'interférence inter-cellulaire représente les interférences dues aux signaux envoyés par les mobiles des cellules voisines et qui viennent constituer du bruit supplémentaire au niveau de la réception à la station de base de la cellule. Dans le sens descendant, l'interférence inter-cellulaire représente les interférences dues aux signaux envoyés par les stations de base des cellules voisines et qui viennent constituer du bruit supplémentaire au niveau du mobile en réception.

1.3.2 Contrôle de puissance en CDMA

Comme nous l'avons vu précédemment, il faut s'assurer qu'à la réception le rapport $\frac{E_b}{N_t}$ est supérieur à un certain seuil critique pour assurer le décodage du signal reçu. Pour cela, le mobile ou la station de base respectivement peut augmenter la puissance qu'il émet. Le but du contrôle de puissance est de veiller à ce que chaque mobile ou chaque station de base envoie le minimum de puissance nécessaire. Le minimum correspond à la puissance nécessaire pour que le rapport $\frac{E_b}{N_t}$ soit au niveau du seuil requis. Ceci améliore la capacité du système en minimisant les interférences. Une autre utilité du contrôle de puissance est de palier aux fluctuations de puissance dues aux pertes dans le milieu extérieur.

Les deux types de contrôle de puissance utilisés dans les systèmes fondés sur le CDMA sont les contrôles de puissance en boucle ouverte (*open-loop*) et en boucle fermée (*closed-loop*). Le contrôle de puissance en boucle ouverte concerne uniquement le mobile (sens montant) et son but est de déterminer le niveau de puissance du signal à transmettre avant de rentrer en communication avec la station de base. Ce niveau de puissance est calculé en fonction de l'affaiblissement de parcours ou *path loss* mesuré dans la voie descendante sur des canaux définis dans ce but par le réseau. Le contrôle de puissance en boucle ouverte permet de compenser des évanouissements à long terme (affaiblissement de parcours dû à des distances importantes entre l'émetteur et le récepteur) et, en particulier, les évanouissements dus au phénomène de l'effet de masque ou *shadowing* (évanouissement dû à la présence d'obstacles tels que des arbres, des collines ou des immeubles).

Une très forte hypothèse qui est faite dans le contrôle de puissance en boucle ouverte est de considérer que les évanouissements dans les voies montante et descendante sont identiques. Or, en mode FDD où la voie descendante et la voie montante se trouvent dans des fréquences différentes, ce type de contrôle de puissance n'est pas très efficace pour compenser les effets des évanouissements rapides car ces derniers dépendent de la fréquence porteuse du signal de transmission et de la vitesse du mobile. En effet, les évanouissements rapides sont caractérisés par des variations rapides de la puissance du signal dans des intervalles de temps assez courts, et ils trouvent leur origine dans les

réflexions du signal transmis sur les différents obstacles et dans la vitesse relative entre le mobile et la station de base. Cela entraîne une dégradation de type “effet doppler” avec un décalage en fréquence.

Le contrôle de puissance en boucle fermée est utilisé dans le but de compenser les évanouissements rapides. Il est appelé “à boucle fermée” puisque, à la différence du contrôle de puissance en boucle ouverte, le récepteur concerné calcule des commandes de contrôle et les envoie à la source émettrice pour que celle-ci règle sa puissance d’émission.

Une fois que le mobile a établi un lien de communication avec la station de base, le contrôle de puissance en boucle fermée est activé. Dans la voie montante, la station de base mesure de manière permanente la qualité du signal en termes du rapport $\frac{E_b}{N_t}$. Si la qualité du signal est en dessus de la valeur $\frac{E_b}{N_t}$ cible, la station de base envoie une commande sur le canal descendant au mobile pour lui demander de réduire la puissance d’émission. Par contre, si la qualité du signal est en dessous de cette valeur, la station de base demande au mobile d’augmenter la puissance d’émission. La technique pour obtenir la valeur du $\frac{E_b}{N_t}$ cible est décrite dans le chapitre 2.

Notons que le contrôle de puissance en boucle fermée a lieu dans les deux sens, c’est-à-dire dans les voies montante et descendante. Ainsi, le terminal mobile, au même titre que la station de base, contrôle la puissance de cette dernière suivant une procédure similaire à celle décrite ci-dessus.

1.3.3 Macrodiversité en CDMA

Dans un réseau cellulaire, la couverture est assurée par des stations de base différentes. Lorsque le mobile est en communication avec une station de base, qu’il se déplace et qu’une autre station de base prend le relais, on dit qu’une procédure de *handover* a lieu. **Soft handover.** En général, pendant la procédure de *handover*, le mobile doit interrompre la communication avec une station de base avant d’en établir une autre avec une station de base différente. C’est le cas dans la plupart des systèmes fondés sur le FDMA et le TDMA. Au contraire, dans un système CDMA où les cellules voisines utilisent la même fréquence porteuse, le mobile peut conserver une liaison radio avec plusieurs stations de base simultanément. Comme il n’y a pas de rupture physique de la communication, ce

type de *handover* est appelé *soft handover*. L'état où un mobile est en liaison avec deux stations de base ou plus est appelé "macrodiversité".

Le *soft handover* se manifeste différemment suivant que l'on considère la voie descendante ou montante, comme le montre la figure 1.7.

Dans la voie descendante (figure 1.7a.), lorsqu'une procédure de *soft handover* a lieu et que le mobile rentre en état de "macrodiversité", le récepteur du terminal mobile reçoit le signal transmis par chaque station de base et les trajets multiples qui l'accompagnent et il les recombine selon le principe du récepteur RAKE (recombinaison des signaux afin d'obtenir un signal unique dont la puissance est renforcée).

Le mécanisme de *soft handover* dans la voie montante est illustré dans la figure 1.7b. où un mobile est simultanément en communication avec deux stations de base. Chaque station de base traite indépendamment le signal reçu et récupère l'information binaire qui est ensuite envoyée au contrôleur des stations de base (BSC) qui va sélectionner l'information avec la meilleure qualité. La qualité est mesurée sur les éventuelles erreurs qui apparaissent sur une trame. Dans le cas où la communication entre le mobile et la station de base A n'est plus acceptable, le contrôleur de stations de base peut supprimer ce lien et ne garder que celui entre le mobile et la station de base B. Toutes ces procédures sont totalement transparentes pour l'utilisateur: à aucun moment la communication n'est interrompue, comme cela est le cas dans d'autres systèmes cellulaires tels que le GSM.

Softer handover. Dans ce cas le mobile établit une communication avec une station de base dans un secteur et, en se déplaçant, passe à un autre secteur, sachant que tous deux sont gérés par la même station de base (la cellule est sectorisée lorsque l'antenne est composée d'antennes directionnelles). Le mobile peut établir simultanément une communication avec deux secteurs gérés par la même station de base lorsqu'il se trouve dans la surface où les deux secteurs se chevauchent. Dans la voie descendante, le mobile combine les signaux provenant des deux secteurs (transmis sur deux canaux distincts par la même station de base) à l'aide d'un récepteur RAKE, comme dans le *soft handover*. En revanche, à la différence du *soft handover*, dans la voie montante, le signal n'est pas combiné au niveau de contrôleur de stations de base, mais directement dans la station de base à l'aide par exemple, d'un récepteur RAKE.

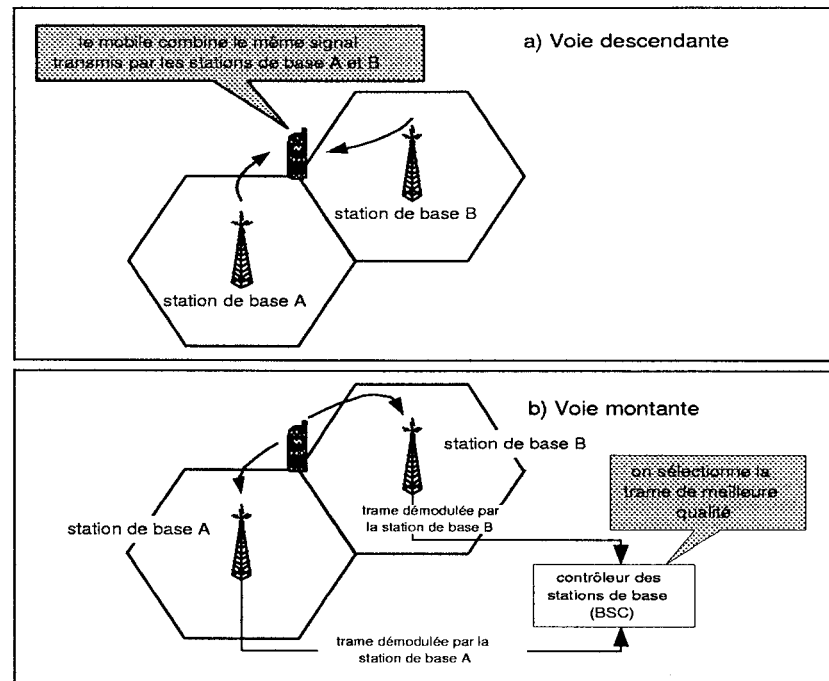


FIG. 1.7 – Principe du soft handover dans un système CDMA

Ainsi l'utilisation du *soft handover* et du *softer handover* permet d'accroître les performances de la liaison en y ajoutant une forme de diversité. Il est cependant nécessaire de minimiser les situations de *soft handover*, sous peine de gaspiller les ressources et diminuer la capacité du système.

1.3.4 Les canaux physiques CDMA2000

1.3.4.1 Canaux physiques sur le lien montant

Sur le lien montant (du mobile vers la station de base), le mobile dispose de plusieurs canaux physiques sur l'interface radio qui lui sont dédiés. Le canal fondamental R-FCH (*Reverse Fundamental Channel*) et les canaux supplémentaires R-SCH (*Reverse Supplemental Channel*) sont utilisés pour transmettre les données de l'utilisateur. Le canal de contrôle dédié R-DCCH (*Reverse Dedicated Control Channel*) transmet les informations de contrôle comme les mesures. Finalement, le canal pilote R-PICH (*Reverse Pilot Channel*) transmet un signal utilisé comme référence pour une détection cohérente des

signaux à la station de base.

1.3.4.2 Canaux physiques sur le lien descendant

Sur le lien descendant (de la station de base vers le mobile), il existe également plusieurs canaux physiques dédiés. Comme pour le sens montant le canal fondamental F-FCH (*Forward Fundamental Channel*) et les canaux supplémentaires F-SCH (*Forward Supplemental Channel*) sont les canaux de transmission des données de l'utilisateur. Le canal de contrôle dédié F-DCCH (*Forward Dedicated Control Channel*) transmet les informations de contrôle. Le canal pilote F-PICH (*Forward Pilot Channel*) transmet un signal utilisé comme référence pour une détection cohérente des signaux. Puis le sens descendant dispose d'autres canaux dédiés par rapport au sens montant: le canal de synchronisation F-SYNCH (*Forward Synchronization Channel*) qui est utilisé par les terminaux mobiles afin de se synchroniser sur un même temps de référence initial, et les canaux de *paging* F-PCH (*Forward Paging Channel*) qui servent à envoyer des messages aux mobiles lorsque ceux-ci ne sont pas encore affectés à un canal de trafic (fondamental ou supplémentaire).

1.3.4.3 Caractéristiques des canaux FCH et SCH

Que ce soit dans le sens montant ou dans le sens descendant, les données sont transmises sur le canal fondamental ou sur un canal supplémentaire. En CDMA2000, sur ces canaux, les données sont transmises sur des trames de longueur 20 ms. De plus, notons que le canal fondamental est utilisé pour la transmission de données à faible débit (voix par exemple) et opère avec un taux d'erreur par trames FER d'environ 1%. Un canal supplémentaire quant à lui, est utilisé pour la transmission de données à plus haut débit et opère avec un taux d'erreur par trames d'environ 5%.

Sur ces canaux de transmissions des données, il existe plusieurs configurations du lien radio, autorisant des débits fixés. Ces configurations sont celles de la norme CDMA2000 et sont référencées dans le standard IS-2000-2^[5] du TTA, qui détaille la couche physique de la norme CDMA2000. Dans le cadre de ce mémoire nous considérons la configuration radio 4 sur le lien descendant, et la configuration radio 3 sur le lien montant, c'est-à-dire que sur chaque lien, montant et descendant, les débits possibles pour la transmission des données

sont les suivants: 9600bps, 19200bps, 38400bps, 76800bps, 153600bps, 307200bps.

Ainsi le signal d'un usager qui transmet ses données sur le lien radio est caractérisé par le FER et le débit du canal qu'il utilise. Par la suite on appelle RAB (*Radio Access Bearer*) la donnée de la valeur du FER et du débit R alloués à une session usager.

1.4 La qualité de service

Depuis quelques années les fournisseurs de service sont confrontés non seulement au problème de croissance explosive du nombre d'utilisateurs mais aussi à des aspects de politique, globalisation et stabilité du réseau. Il s'agit maintenant d'offrir aux utilisateurs une qualité de service constante et de donner des priorités à certains types de trafic. De plus, l'Internet actuel est un réseau à commutation de paquets à l'architecture et aux principes simples, qui ne fournit qu'un service dit "au mieux" (BE pour *Best Effort*) c'est-à-dire sans aucune garantie sur le débit disponible, le délai subi ou encore les pertes de paquets. Ce service s'avère suffisant pour les applications "élastiques" de transfert de données mais insuffisant pour les nouvelles applications multimédias qui se développent et dont les contraintes temps réel se satisfont difficilement du service au mieux. La Qualité de Service (QoS) englobe ainsi tous les mécanismes permettant de différencier les types de trafic, ceux-ci pouvant être classés et administrés différemment à travers le réseau.

1.4.1 Critères de qualité de service

Avant de parler de Qualité de Service, nous allons décrire les critères qui caractérisent le service dans le réseau permettant de délivrer un service de bonne qualité. La caractérisation de la qualité de service est généralement exprimée par les critères suivants.

- **Délai:** temps écoulé entre l'envoi d'un paquet par un émetteur et sa réception par le destinataire. Le délai tient compte du délai de propagation le long du chemin et du délai de transmission induit par la mise en file d'attente des paquets dans les systèmes intermédiaires.
- **Gigue:** variation du délai de bout en bout.
- **Bande passante ou débit maximum:** taux de transfert maximum pouvant être

maintenu entre deux points terminaux.

- **Fiabilité:** taux moyen d’erreurs d’une liaison.

On dira qu’un réseau peut fournir de la qualité de service s’il est capable de fournir un service adapté aux besoins spécifiques des applications (en particulier temps réel): garantie de débit ou de délai par exemple.

1.4.2 Classes de service différenciées et priorités d’usager

1.4.2.1 Les différentes classes de qualité de service

La QoS est généralement assimilée à la discrimination des services, autrement dit à la définition de classes différenciées de services. Le groupe 3GPP^[6] prévoit différents services dans les futurs systèmes de troisième génération susceptibles d’être classés suivant quatre catégories:

- la classe conversationnelle (Co),
- la classe diffusion (St),
- la classe interactive (Int),
- la classe tâche de fond (Bg).

La différence entre ces classes de service vient de la sensibilité de chacune par rapport au délai: la classe conversationnelle englobe les applications très sensibles au délai tandis que la classe tâche de fond englobe les applications les moins sensibles au délai. Ainsi les classes conversationnelle et diffusion regroupent les applications avec des contraintes temps réel, la différence entre ces deux classes venant uniquement de leur degré de sensibilité au délai. Les classes interactive et tâche de fond regroupent les applications traditionnelles de l’internet comme la navigation sur l’Internet, le courrier électronique, le transfert de fichiers. Ces applications sont beaucoup moins sensibles au délai que les classes conversationnelle et diffusion mais sont en revanche beaucoup plus sensibles aux erreurs de transfert. La principale différence entre les classes interactive et tâche de fond est que les applications de type interactive sont des applications interactives comme la messagerie instantanée ou la navigation interactive sur l’Internet, tandis que les applications de

type tâche de fond sont des applications en tâche de fond comme le téléchargement de courriels en tâche de fond ou le téléchargement de fichiers en tâche de fond.

La classe conversationnelle (*conversational*). Les applications de cette classe nécessitent un service bidirectionnel en temps réel impliquant deux utilisateurs humains ou plus. Les contraintes dépendent donc de la perception humaine: la limite sur le délai maximum toléré est une limite stricte car toute dégradation sur le délai induirait une perte de qualité notable dans la perception humaine du signal. Les exemples de ce type d'applications sont la téléphonie, la vidéophonie, la voix sur IP, les jeux interactifs.

La classe diffusion (*streaming*). Les applications de cette classe impliquent un utilisateur humain et un serveur de données. Ce sont des applications temps réel asymétriques où les données sont transférées du réseau vers les mobiles. Le manque d'interactivité entre l'utilisateur et la source de données autorise des délais un peu plus importants que dans les cas des applications de type conversationnelle, et ce sans perturber la QoS. Les exemples d'applications de type diffusion sont les nouvelles applications issues de l'Internet, telles que les applications audio ou vidéo sur demande.

La classe interactive (*interactive*). Les applications de cette classe impliquent un utilisateur (machine ou humain) dialoguant avec un serveur de données ou d'applications. Contrairement aux deux classes précédentes, les performances temps réel ne sont pas nécessaires, il s'agit seulement d'attendre un certain temps pour répondre aux requêtes. Par contre les informations ne doivent pas être altérées. Les exemples d'applications de type interactive sont la navigation sur l'Internet, l'accès aux bases de données ainsi qu'aux serveurs d'applications.

La classe tâche de fond (*background*). Les applications de cette classe impliquent un utilisateur, le plus souvent un équipement terminal, réalisant l'envoi et la réception de données en tâche de fond. L'absence d'interactivité pour ces applications fait que l'utilisateur à l'origine de la requête n'est pas en attente d'une réponse dans une limite de temps fixée. Ce sont donc les applications les moins sensibles au délai, mais sont très sensibles aux erreurs sur l'information transférées. Les exemples d'applications de type tâche de fond sont le courrier électronique, le transfert de messages courts (SMS pour *Short Messages Services*), le téléchargement de données ou de fichiers.

Le tableau 1.3 récapitule les applications susceptibles d'être offertes sur les systèmes de troisième génération suivant le groupe 3GPP^[6].

TAB. 1.3 – *Les 4 classes de service définies par 3GPP*

Classes de service	Caractéristiques	Exemples	Exigence en QoS
Conversationnelle	- transport bidirectionnel entre deux personnes - contraintes de QoS basées sur la perception humaine - conserve la relation temporelle entre les entités du flot - temps réel	voix sur IP jeux vidéos vidéophonie	faible délai faible gigue
Diffusion	- transport unidirectionnel vers une personne - conserve la relation temporelle entre les entités du flot - temps réel	vidéo sur demande audio sur demande multimédia	faible délai
Interactive	- transport bidirectionnel entre un usager et un serveur - conserve la charge utile - non temps réel	surf sur internet accès bases de données	faible FER
Tâche de fond	- transport unidirectionnel vers une machine - conserve la charge utile - non temps réel	courriel, SMS transfert de fichiers transfert de données	faible FER

1.4.2.2 Profils d'utilisateur

En plus de différencier les services offerts aux usagers, il est également nécessaire de classer entre eux les usagers utilisant un même type d'application afin de construire une hiérarchie dont les critères sont choisis par l'opérateur, comme la tarification par exemple. Ainsi nous utiliserons les applications types mentionnées dans la table 1.3 et la classification hiérarchique des usagers établie par Troeung^[7] et Jaumard *et al.*^[8]: les usagers sont classés suivant quatre catégories précisées par ordre décroissant de priorité: Or (Go pour *Gold*), Argent (Si pour *Silver*), Bronze (Br pour *Bronze*) et Fer (Ir pour *Iron*). Les critères de cette classification sont laissés à la discrétion de l'opérateur. Par l'appartenance à l'une de ces catégories, l'utilisateur bénéficie d'une priorité relative lors de l'accès au réseau ainsi que d'une garantie plus ou moins fiable de la qualité de service fournie. Ainsi, pour une même application mise en œuvre, un abonné de la classe Go reçoit un niveau de QoS supérieur à un usager de la classe Ir ainsi qu'un taux de blocage (NdS pour Niveau de Service) inférieur à l'entrée du réseau.

Dans le cadre de l'étude, la donnée d'une instance d'application aux caractéristiques spécifiées par la table 1.3 et d'une priorité de l'utilisateur désigne une session. Ces deux critères permettent donc de prioriser les différentes sessions suivant le type d'application mise en œuvre et la priorité de l'utilisateur au service.

Le tableau 1.4 indique la hiérarchisation choisie pour les sessions d'abonnés supportées par le réseau de troisième génération.

TAB. 1.4 – *Hiérarchisation relative des sessions suivant le type d'application et la priorité de l'utilisateur*

	Go	Si	Br	Ir
Co	15	14	13	12
St	11	10	9	8
Int	7	6	5	4
Bg	3	2	1	0

1.4.3 Modèle à intégration de services pour le réseau cœur

Le groupe de travail IntServ (pour *Integrated Services*) de l'IETF^[9] (pour *Internet Engineering Task Force*) a été créé en 1994 pour définir un modèle de service de l'Internet amélioré, capable de transporter aussi bien de l'audio, de la vidéo, des données temps réel que du trafic de données classiques; soit transformer l'Internet en un réseau à intégration de services.

L'architecture globale du modèle est la suivante. La qualité du service fourni par le réseau à un flot dépend des ressources (bande passante et tampons mémoire) disponibles. Pour garantir une qualité de service à un flot, il faut s'assurer qu'une quantité de ressources suffisante est disponible dans chacun des routeurs (et liens) qu'il traverse pendant toute la durée du flot. Comme le montre la figure 1.8, dans chaque routeur on appelle "contrôle de trafic" l'ensemble des éléments qui permettent d'assurer cette condition, au nombre de trois:

- le classifieur de paquets détermine à quel flot appartient un paquet,
- l'ordonnanceur de paquets sélectionne pour émission les paquets des différents flots en fonction des ressources réservées,
- le contrôle d'admission vérifie que suffisamment de ressources sont disponibles au moment de l'établissement de la réservation.

Le quatrième élément fondamental d'un réseau à intégration de services est le protocole de réservation des ressources, qui à partir de la requête formulée par une application

fournit les informations au contrôle de trafic de chaque routeur appartenant au chemin du flot et récupère la réponse du module de contrôle d'admission. Le protocole choisi pour IntServ et que nous utilisons dans le cadre de notre travail est RSVP^[10] (pour *Resource reSerVation setup Protocol*).

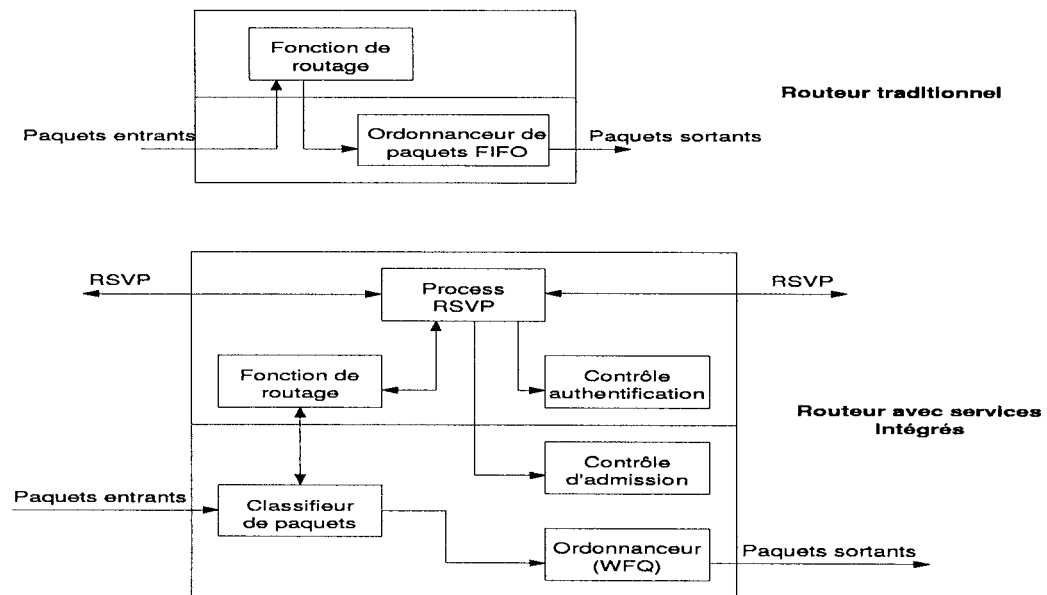


FIG. 1.8 – Architecture d'un nœud Integrated Services

Nous décrivons maintenant brièvement les mécanismes de RSVP. Dans l'architecture Intserv, chaque flot est caractérisé par un TSPEC (pour *Traffic Specification*) qui comprend les paramètres de spécification du trafic.

Une session RSVP est identifiée par une adresse IP destination (éventuellement multi-cast) et un numéro de session. L'application émettrice d'une session envoie des messages PATH contenant le TSPEC du flot qui sont routés jusqu'au destinataire. Dans chacun des routeurs qu'ils traversent, ces messages PATH créent un *Path State* contenant essentiellement l'adresse du routeur précédent, créant ainsi un arbre de parcours inverse. Un appel à la fonction de routage permet ensuite de déterminer le lien de sortie à utiliser pour envoyer le message PATH vers le nœud suivant.

Quand l'application réceptrice reçoit le message PATH, elle connaît alors les caractéristiques du flot de données proposé et celles du chemin parcouru. Elle spécifie la qualité

de service qu'elle désire obtenir en envoyant un message RESV qui "remontera" vers l'émetteur grâce aux *Path States* installés par le message PATH. Ce message RESV contient le "TSPEC du récepteur" caractérisant le trafic pour lequel le récepteur veut réserver des ressources et un RSPEC (*Resource Specification*) spécifiant les ressources à réserver. A chaque nœud, un contrôle d'admission est réalisé pour vérifier que suffisamment de ressources sont disponibles pour fournir le service demandé. Si c'est le cas, un *Resv State* est créé, contenant les informations relatives à la réservation à effectuer et le message est envoyé au nœud amont. Les états de classification et d'ordonnancement peuvent alors être mis en place.

Les états maintenus dans les routeurs étant "relâchés", les messages sont réémis périodiquement pour les rafraîchir, sinon ils disparaissent.

La figure 1.9, tirée de la documentation du logiciel OPNET, représente un réseau mettant en œuvre RSVP.

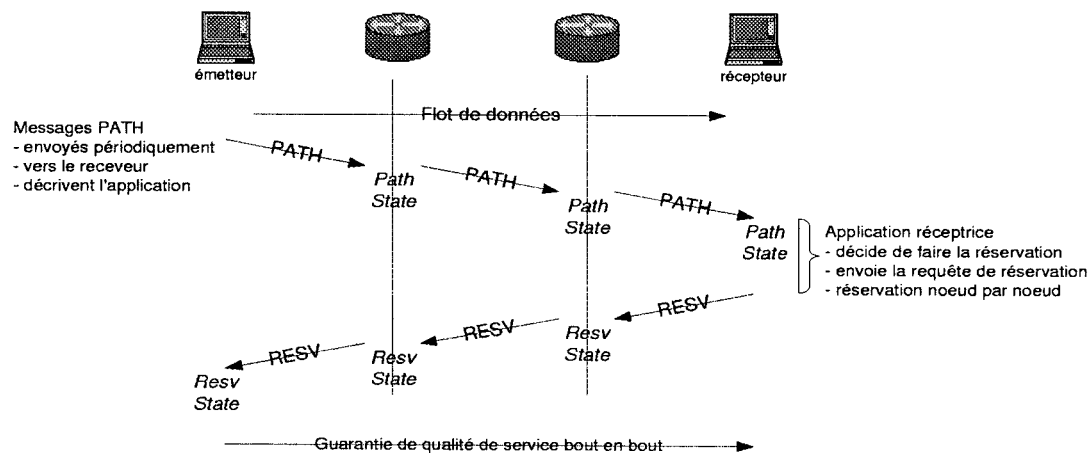


FIG. 1.9 – Protocole de réservation RSVP

En résumé, RSVP possède les attributs suivants:

- RSVP est unidirectionnel, car il n'effectue les réservations que pour les flux de données unidirectionnels,
- RSVP est orienté receveur, car le receveur d'un flux de données initie et maintient la réservation des ressources utilisées pour ce flux,
- RSVP maintient des états souples dans les routeurs et les hôtes,

- RSVP n'est pas un protocole de routage mais dépend fortement du protocole utilisé,
- RSVP transporte et maintient les paramètres de contrôle de trafic et de contrôle de politique qui lui sont opaques,
- RSVP fournit des opérations transparentes à travers les routeurs qui ne le supportent pas.

1.4.4 Implémenter la qualité de service

Pour implémenter des classes de service et garantir des ressources, il faut définir une ou plusieurs politiques sur les nœuds du réseau permettant d'implémenter la Qualité de Service demandée, en utilisant des mécanismes de contrôle d'admission. Or l'acceptation ou le rejet de sessions dans le réseau dépend de l'ensemble des politiques de contrôle adoptées, parmi lesquelles la priorisation, décrite dans la section 1.4.2, et la préemption du trafic dans le réseau. Nous allons ici décrire les différents mécanismes de préemption du trafic, c'est-à-dire les différentes techniques de gestion des files d'attente (ordonnancement des paquets), qui ont un rôle important dans la différenciation et la garantie des services, ainsi que les techniques de lissage du trafic (*traffic shaping*) qui permettent de contrôler le volume et le débit du trafic entrant dans le réseau.

La stratégie de gestion des diverses files d'attente sur un routeur joue en effet un rôle essentiel dans la différenciation des services, à travers le choix de l'algorithme qui place les paquets dans la queue de sortie et le choix de la taille maximale de la queue.

Un grand nombre de disciplines de service ont été développées qui permettent d'offrir des garanties de service plus ou moins complètes. Une analyse et une comparaison de ces disciplines sont présentées dans les articles de Kurose^[11] et Aras *et al.*^[12]. Dans le cadre de notre travail, nous avons utilisé la discipline d'attente équitable proportionnelle car elle semble particulièrement prometteuse.

1.4.4.1 Premier arrivé, Premier sorti (FIFO pour *First In, First Out*)

Il s'agit de la méthode standard de gestion de trafic entre une interface d'entrée et une interface de sortie. Les paquets sont placés dans la file de sortie dans l'ordre dans

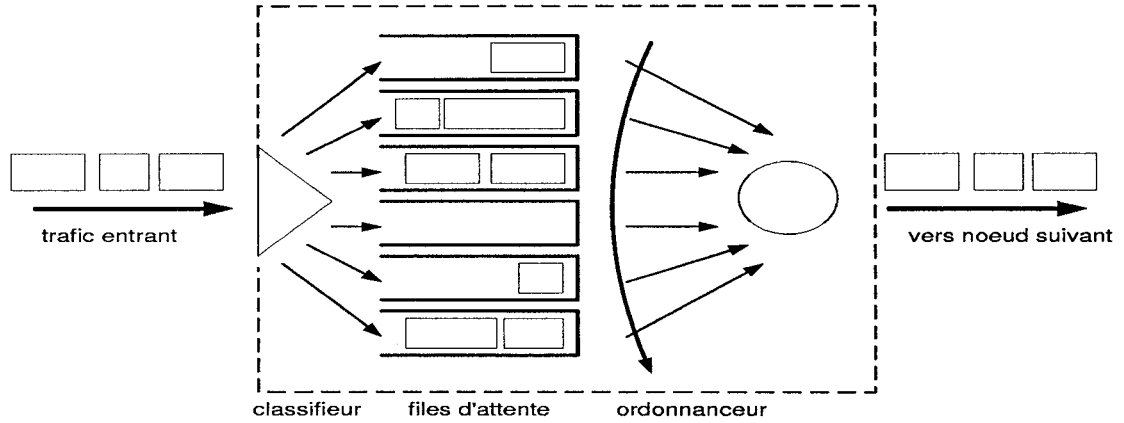
lequel ils sont reçus. Cette discipline, utilisée dans les routeurs Internet actuels, permet une allocation des ressources du routeur flexible et simple. De plus, elle traite les paquets indépendamment les uns des autres et ne nécessite pas de garder un “état” dans le réseau. Malheureusement, elle ne permet pas d’isoler les utilisateurs “mauvais joueurs”, c’est-à-dire qui monopolisent les ressources du réseau en refusant par exemple de s’adapter à la congestion dans le réseau, des utilisateurs “bon joueurs”.

1.4.4.2 Attente Equitable

L’algorithme d’attente équitable (FQ pour *Fair Queueing*) est une idée originale de Nagle^[13] développée ensuite de façon substantielle par Demers *et al.*^[14]. Considérons un routeur et une liaison partagée par plusieurs connexions. L’idée de l’algorithme FQ est d’isoler les connexions afin que les comportements anti sociaux d’une ou de plusieurs de ces connexions n’aient pas d’impact sur les connexions raisonnables. Le principe est de séparer le trafic arrivant dans un routeur en plusieurs flots, et d’allouer à chaque flot une fraction de la bande passante. Le schéma d’un tel type d’ordonnanceur est donné à la figure 1.10.

L’allocation de bande passante peut être réalisée de façon relativement simple, en servant les paquets en tête de chaque file (et donc de chaque flot) à tour de rôle. Si une file est vide, on passe son tour et on sert la file suivante. Cet algorithme a cependant des limitations. En particulier, il ne tient pas compte de la taille des paquets. Ainsi un flot dans lequel la taille des paquets est élevée utilisera en fait une bande passante plus élevée qu’un flot dans lequel la taille des paquets est petite. De plus, l’algorithme pénalise les paquets d’un flot qui arrivent juste après que la file d’attente de ce flot ait été examinée par le routeur. Ces paquets devront attendre que toutes les autres files soient examinées et servies avant d’être considérés. De façon plus générale, l’algorithme favorise les sources dont le trafic a le plus de variabilité.

Pour remédier à ces problèmes, Demers *et al.*^[14] ont développé une nouvelle version de l’algorithme d’attente équitable. Considérons d’abord une approximation “fluide” de la discipline FQ dans laquelle les files sont servies chacune à leur tour, et un bit seulement de chaque file est envoyé lors de chaque service. Cette discipline appelée “à tour de rôle

FIG. 1.10 – *Ordonnanceur à isolation de flux*

bit par bit" (BBRR pour *bit-by-bit round-robin*), est plus équitable que celle de Nagle^[13] car elle prend en compte la taille des paquets, et le bit en tête d'une file attend avant d'être servi au plus $n - 1$ fois le temps de transmission d'un bit si n est le nombre de files (et donc de flots) actives.

Etant donné un débit ou une capacité disponible r et un nombre n de connexions actives, la fraction de bande passante allouée à chaque flot est $\frac{1}{n}.r$. De plus, il est possible de calculer le moment auquel le dernier bit d'un paquet de longueur L aurait été transmis selon la discipline BBRR. L'idée de l'algorithme d'attente équitable est justement de transmettre les paquets selon un ordre correspondant au temps auquel leur dernier bit aurait été envoyé selon la discipline BBRR. On exprime cela rigoureusement de la façon suivante.

Soit $R(t)$ le nombre de tours effectués par l'algorithme BBRR à l'instant t . Ainsi un paquet de taille L bits dont le premier bit est servi à l'instant t_0 aura son dernier bit servi L tours plus tard, soit au temps t tel que $R(t) = R(t_0) + L$. Soient t_i^k le temps d'arrivée du paquet k du flot i , L_i^k la taille de ce paquet, S_i^k la valeur de $R(t)$ quand le paquet commence son service, et F_i^k la valeur de $R(t)$ quand le paquet termine son service. On a donc

$$F_i^k = S_i^k + L_i^k,$$

$$S_i^k = \max \left(F_i^{k-1}, R(t_i^k) \right).$$

La discipline FQ envoie alors les paquets dans l'ordre croissant des F_i^k , c'est-à-dire que lorsqu'un paquet vient d'être transmis, le paquet suivant à transmettre est le paquet avec la plus petite valeur de F_i^k .

De façon plus générale, on peut associer un poids ϕ_i à un flot i , qui reçoit alors une fraction $\phi_i / \sum_j \phi_j$ de la bande passante. On obtient alors la discipline *Weighted Fair Queueing*, qui a plus tard été réinventée par Parekh et Gallager^{[15][16]} sous le nom de PGPS. Comme WFQ et PGPS sont le même algorithme, nous reportons la description de l'algorithme complet au paragraphe consacré à PGPS.

1.4.4.3 *Generalized Processor Sharing*

Parekh et Gallager^[15] ont défini la discipline de service *Generalized Processor Sharing* (GPS), un modèle théorique de multiplexage de flots (ou sessions) de la façon suivante. Un serveur GPS conserve le travail et opère à un débit fixé r . Il est caractérisé par des nombres réels positifs $\phi_1, \phi_2, \dots, \phi_N$. Soit $S_i(\tau, t)$ la quantité de service reçue par la session i dans l'intervalle de temps $[\tau, t]$. Le serveur GPS est défini tel que

$$\frac{S_i(\tau, t)}{S_j(\tau, t)} \geq \frac{\phi_i}{\phi_j}, \quad j = 1, 2, \dots, N$$

pour toutes les sessions i actives (i.e. qui ont du trafic en attente) sur tout l'intervalle de temps $[\tau, t]$.

En sommant

$$S_i(\tau, t) \phi_j \geq \phi_i S_j(\tau, t)$$

pour tous les j on obtient

$$S_i(\tau, t) \sum_j \phi_j \geq \phi_i \sum_j S_j(\tau, t),$$

$$S_i(\tau, t) \geq \frac{\phi_i}{\sum_j \phi_j} r(t - \tau).$$

En conséquence, une part de la bande passante est garantie à la session i :

$$g_i = \frac{\phi_i}{\sum_j \phi_j} r.$$

Si l'on assure que $\sum_j \phi_j \leq 1$, cette garantie s'exprime alors en termes de bande passante absolue.

GPS est une discipline idéalisée qui suppose que le serveur peut servir plusieurs sessions simultanément. Il est important de noter que si cela permet de qualifier ce modèle de "fluide", il manipule cependant des paquets. En particulier le temps de traversée du serveur par un paquet de taille L_i appartenant à la session i est égal à sa taille divisée par la bande passante allouée à la session i , soit son délai de paquetsation L_i/g_i .

1.4.4.4 *Packet by Packet* GPS ou Attente équitable proportionnelle (WFQ pour *Weighted Fair Queuing*)

La discipline de service d'attente équitable proportionnelle a suscité beaucoup de travaux depuis de nombreuses années. La raison de sa popularité tient sans doute à sa flexibilité: elle peut être utilisée pour le contrôle de congestion (par partage équitable de la bande passante entre tous les flots), pour fournir des garanties de débit, de délai (comme cela est montré en 1.4.5), pour faire du partage de lien (entre organisations, classes de trafic) et, pour finir, elle peut combiner (de façon plus ou moins efficace selon les variantes) toutes ces fonctionnalités de manière hiérarchique pour un modèle à intégrations de service complet.

Parekh et Gallager^[15] proposent une émulation de GPS pour des réseaux réels, qui ne peuvent transmettre qu'un paquet à la fois. Ils l'appellent PGPS pour *Packet by Packet* GPS. L'idée est de servir les paquets en les triant sur leur date de fin de transmission sous GPS, c'est-à-dire que les paquets seront émis dans l'ordre où ils finissent (et non pas commencent) leur transmission avec un serveur GPS.

Ils montrent que le retard de service d'un paquet k sous PGPS par rapport à GPS est

borné par le temps de transmission d'un paquet de taille maximum:

$$F_{PGPS}^k - F_{GPS}^k \leq \frac{L_{\max}}{r}.$$

Il s'agit donc de trier les paquets avec une estampille correspondant à la date de leur fin de transmission en GPS exprimée en "temps virtuel". Ce temps virtuel $V(t)$ est défini de façon que le service fourni à un flot actif soit constant dans le référentiel de ce temps virtuel. Cela signifie par exemple que si tous les flots sont actifs, le temps virtuel progresse à la même vitesse que le temps réel; si la moitié seulement des flots sont actifs, ils reçoivent un service double, le temps virtuel progresse deux fois plus vite que le temps réel et le service en temps virtuel reste donc constant.

Posons que le $k^{\text{ème}}$ paquet de la session i arrive au temps a_i^k et a pour taille L_i^k . Appelons S_i^k et F_i^k les dates respectives en temps virtuel auxquelles le paquet commence et termine son service avec le serveur GPS (F_i^0 étant initialisé à zéro), on a:

$$S_i^k = \max \left(F_i^{k-1}, V(a_i^k) \right),$$

$$F_i^k = S_i^k + \frac{L_i^k}{\phi_i}.$$

L'opération $\max$ pour le S_i^k assure que le paquet ne commence pas son service avant que le paquet précédent du même flot n'ait terminé sa transmission (auquel cas le serveur GPS traiterai simultanément deux paquets du même flot). Le calcul du temps virtuel $V(a_i^k)$ pose un problème de complexité car il doit être calculé pour chaque paquet et son calcul est assez complexe.

Comme dit précédemment, PGPS et WFQ sont le même algorithme approximant GPS. Les auteurs du WFQ le concevaient comme une émulation d'un tourniquet bit à bit, aussi la notion de temps virtuel était remplacée par le numéro du cycle de service (comme décrit ci-dessus pour l'algorithme FQ). D'autres approximations du GPS existent, comme SCFQ^[17] pour *Self-Clocked Fair Queuing*, WF²Q^[18] pour *Worst-case Fair Weighted Fair Queuing*, SFQ^[19] pour *Start Time Fair Queuing*, et VC^[20] pour *Virtual Clock*.

1.4.4.5 *Traffic Shaping*

En alternative ou en complément à ces techniques de gestion des files d'attente autres que FIFO, d'autres techniques sont utilisées, telles que le *traffic shaping* qui permet de contrôler le volume de trafic entrant dans le réseau ainsi que le débit avec lequel il est transmis. Les deux techniques principales de *traffic shaping* sont les suivantes.

– Sceau percé ou *Leaky bucket*

Le trafic arrivant dans la file d'attente est régulé pour sortir en flux fixes sur le réseau. La taille de la file d'attente et le taux de transmission sont configurables pas l'utilisateur. Ce mécanisme est bien adapté pour envoyer des débits fixes sur le réseau mais, en contrepartie, ne permet pas d'utiliser plus de ressources lorsque le réseau est peu chargé.

– *Token bucket*

Le trafic ne traverse pas effectivement le sceau, mais est transmis sur la base de jetons présents dans le sceau. Il peut être représenté par un réservoir de capacité σ jetons se remplissant au rythme de ρ jetons/s, comme cela est montré à la figure 1.11. Si le réservoir est plein, les jetons qui arrivent sont perdus. Quand un paquet est émis sur le réseau, il consomme sa taille (en-tête compris) en jetons.

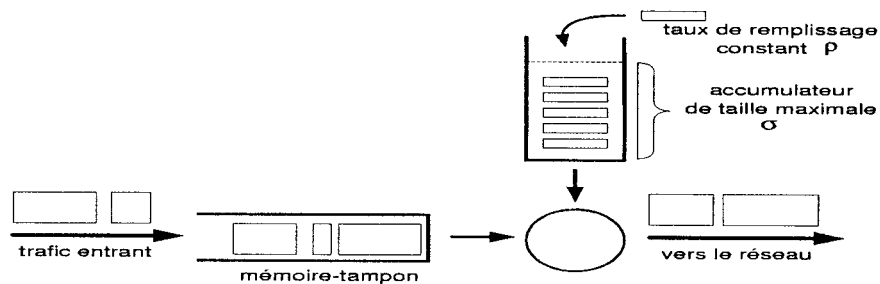


FIG. 1.11 – *Token Bucket*

Un trafic est dit conforme au *token bucket* si lors de l'émission des paquets, il y a assez de jetons dans le réservoir. La figure 1.12 donne la caractérisation d'un trafic en référence à un *token bucket*. Pendant la période:

1. la source émet des données au débit ρ , les jetons sont consommés au rythme

de leur arrivée,

2. le débit de la source augmente, le réservoir de jeton se vide, mais le trafic est toujours conforme au *token bucket*,
3. le débit est inférieur au débit de remplissage du réservoir, celui-ci se remplit,
4. le débit de la source augmente, les capacités du réservoir sont consommées. Le réservoir est vide, le trafic en excédant n'est pas conforme au *token bucket*,
5. la source réduit son débit, le réservoir se remplit à nouveau, puis le débit restant inférieur à ρ , le réservoir est plein, les jetons qui arrivent sont perdus,
6. la source augmente son débit.

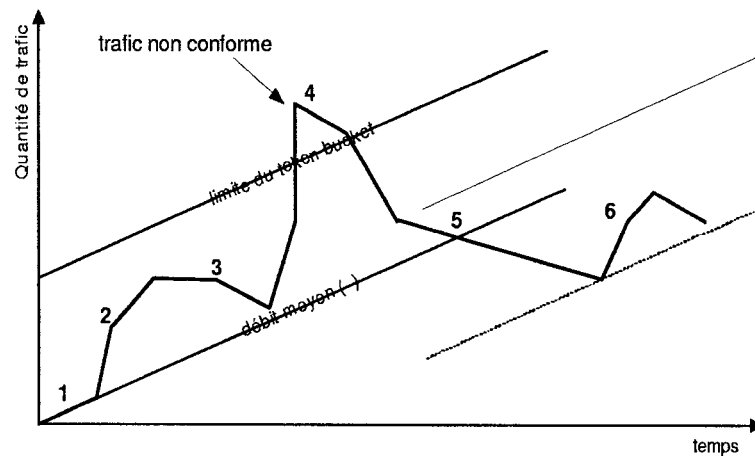


FIG. 1.12 – Caractérisation d'un trafic

Un *token bucket* est défini à l'aide des paramètres σ et ρ , mais aussi à l'aide du débit crête p qui a pour but de limiter la mémoire dans les files d'attente en fournissant une borne maximum plus stricte.

1.4.5 Garantie de délai

1.4.5.1 Courbes d'arrivée et de service

La courbe d'arrivée maximale (dans un modèle abstrait fluide) d'un flot caractérisé par les paramètres (σ, ρ, p) de son *token bucket* et ayant $L_{\max}$ comme taille maximum

de paquet est, selon Parekh et Gallager^[15]:

$$A(\tau) = \min \{p\tau, \sigma + \rho\tau\}$$

ce qui s'explique de la manière suivante. A long terme (τ grand) le service maximum est obtenu quand le flot a émis une rafale de taille maximum (σ) et continue d'émettre à son débit moyen (ρ). Ceci est exprimé par le terme $\sigma + \rho\tau$. Cependant, la rafale ne peut pas être émise instantanément, le débit étant limité par le débit de crête (p). Ainsi, si à partir de $\tau = 0$ la source émet une rafale, elle pourra émettre à son débit crête jusqu'à ce que la taille maximum de rafale ait été atteinte, après quoi elle devra se limiter à son débit moyen. Donc, pour τ petit, la courbe serait $p\tau$.

Il faut cependant tenir compte du fait que le flot n'est pas fluide mais constitué de paquets. Quel que soit le lissage mis en œuvre, un paquet est toujours émis au débit du lien, que l'on considère comme beaucoup plus important que le débit crête du flot. Ainsi la courbe d'arrivée du flot pour des valeurs de τ petites devient $L_{\max} + p\tau$, ce qui correspond à une augmentation de l'effet des rafales.

$$A(\tau) = \min \{L_{\max} + p\tau, \sigma + \rho\tau\}.$$

Considérons maintenant la courbe de service minimale garantie par le réseau (par une discipline d'ordonnancement comme WFQ par exemple). On a un débit g assuré par le réseau au travers de la discipline d'ordonnancement. Puis il faut prendre en compte les termes d'erreurs de chaque nœud. En effet dans chaque nœud il faut tenir compte de la déviation par rapport au modèle fluide de l'ordonnanceur mis en œuvre. Notons E_m ce terme d'erreur (correspondant à un délai) dans un nœud m et considérons que l'ensemble du réseau ne forme qu'un seul système caractérisé par un terme d'erreur $E_{tot} = \sum E_m$.

La figure 1.13 montre comment ces deux courbes peuvent être utilisées pour calculer le délai $D_{\max}$ maximum subi par un paquet (délai d'attente dans les queues) et la taille maximum $Q_{\max}$ des tampons nécessaires dans un nœud du réseau, dont les valeurs sont données explicitement dans le cas de WFQ dans la section 1.4.5.2.

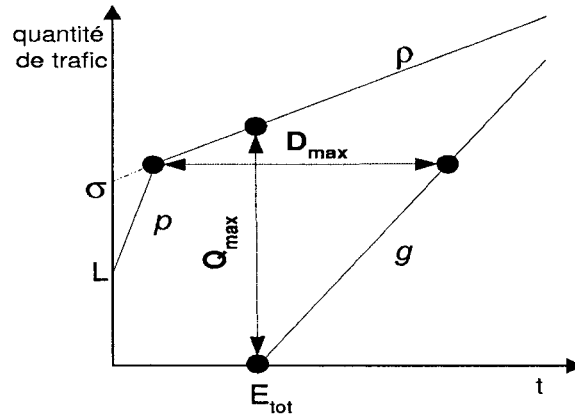


FIG. 1.13 – Courbes d'arrivée et de service

1.4.5.2 Garantie de délai sous la discipline WFQ

La classe de service garanti est conçue pour fournir aux applications une garantie stricte sur le délai subi par chaque paquet d'un flot pendant qu'il traverse les nœuds du réseau. Ce délai est composé de deux parts, le temps de propagation lié au support physique et incompressible, et le temps passé dans les files d'attente, que l'on peut contrôler par divers mécanismes. Nous nous intéressons ici uniquement au temps passé dans les files d'attente, le temps de propagation pouvant être facilement calculé et ajouté pour obtenir un délai total.

Puisque l'algorithme d'attente équitable proportionnelle isole les flots les uns des autres dans les routeurs, on peut s'attendre à ce qu'il permette d'offrir des garanties de performance aux utilisateurs du réseau. C'est effectivement le cas si le trafic généré par les utilisateurs n'est pas trop chaotique (ce point est discuté plus en détail ci dessous). Notons tout d'abord que l'algorithme alloue sur chaque nœud une fraction de la bande passante disponible r à chacun des flots

$$g_i = \frac{\phi_i}{\sum_j \phi_j} r$$

où ϕ_i est le poids associé au flot i . Cette fraction peut être arbitrairement petite si le nombre de connections actives est grand. Cependant, chaque flot est assuré d'une fraction

minimum de capacité.

Les travaux de Parekh et Gallager^[16] ont montré que l'algorithme d'attente équitable proportionnelle permet d'offrir en plus de cette garantie de bande passante une garantie de délai si le trafic généré par les utilisateurs n'est pas trop chaotique. Il faut alors supposer que le trafic de chaque utilisateur est passé à travers un mécanisme de contrôle d'accès de type *token bucket*. Le mécanisme est donc complètement caractérisé par ses paramètres σ et ρ . Il limite le débit moyen de la source (qui est égal à ρ) et le nombre maximal de paquets qui peuvent être émis en rafale (qui est égal à σ paquets).

Considérons maintenant un flot i qui est passé à travers un mécanisme de contrôle de type "seau percé" de paramètres σ_i et ρ_i avant de rentrer dans le réseau. On sait alors qu'à chaque nœud m sur le chemin du flot, le débit de service minimum garanti est:

$$g_i^m = \frac{\phi_i^m}{\sum_j \phi_j^m} r^m$$

où r^m est la bande passante disponible au nœud m et ϕ_i^m le poids de la session i au nœud m .

Ainsi le taux de service minimum sur le chemin est:

$$g_i = \min_m g_i^m.$$

Dans le cadre de l'étude, nous énonçons les résultats établis par Pareck et Gallager dans le cas des systèmes stables, c'est-à-dire pour les systèmes où

$$\sum_j \rho_j \leq r^m \tag{1.1}$$

ce qui peut être assuré par un contrôle d'admission.

Si la session i est localement stable c'est-à-dire si $g_i \geq \rho_i$, alors les travaux de Parekh et Gallager^[16] montrent que:

- la quantité de trafic en attente de service maximale à chaque nœud m sur le chemin ne

dépasse pas une valeur maximum $Q_i^{\max}$ égale à:

$$\begin{aligned} Q_i^{\max} &= \sigma_i && \text{en version GPS,} \\ Q_i^{\max} &= \sigma_i + L_i^{\max} && \text{en version PGPS,} \end{aligned}$$

- le délai maximal des paquets de ce flot entre une source et une destination à travers un réseau de topologie arbitraire ne dépasse pas une valeur maximum $D_i^{\max}$ qui est égale à:

$$\begin{aligned} D_i^{\max} &= \frac{\sigma_i}{g_i} && \text{en version GPS,} \\ D_i^{\max} &= \frac{\sigma_i}{g_i} + \frac{2(h_i - 1)L_i^{\max}}{g_i} + \sum_{m=1}^{h_i} \frac{L_{\max}}{r^m} && \text{en version PGPS,} \end{aligned}$$

où r^m est la bande passante disponible au nœud m , h_i est le nombre de nœud entre la source et la destination, $L_i^{\max}$ la taille maximale des paquets du flot i et $L_{\max}$ la taille maximale d'un paquet dans le réseau.

Soit la condition RPPS (*Rate Proportional Processor Sharing*) $\phi_i^m = \rho_i$ à chaque nœud m , alors la session i est localement stable. De plus, de cette manière, les coefficients ϕ_i affectés aux sessions i sont alors en adéquation avec les qualités de service demandées par les sessions puisque ρ_i représente le débit régulant l'émission de la source. Enfin, les travaux de Parekh et Gallager^[16] montrent que sous la condition RPPS:

$$\begin{aligned} Q_i^{\max} &= \sigma_i && \text{en version GPS,} \\ Q_i^{\max} &= \sigma_i + L_i^{\max} && \text{en version PGPS,} \end{aligned} \quad (1.2)$$

$$\begin{aligned} D_i^{\max} &= \frac{\sigma_i}{\rho_i} && \text{en version GPS,} \\ D_i^{\max} &= \frac{\sigma_i}{\rho_i} + \frac{2(h_i - 1)L_i^{\max}}{\rho_i} + \sum_{m=1}^{h_i} \frac{L_{\max}}{r^m} && \text{en version PGPS.} \end{aligned} \quad (1.3)$$

Pour cette dernière formule en version PGPS le premier terme est souvent le terme prépondérant, en particulier si l'on considère un réseau haut débit car r^m devient très grand devant $L_{\max}$ et l'on obtient alors la formule:

$$D_i^{\max} = \frac{\sigma_i}{\rho_i} + \frac{2(h_i - 1)L_i^{\max}}{\rho_i} \quad \text{en version PGPS}$$

On s'aperçoit que l'on retombe sur la formule en version GPS si $L_i^{\max}$ est très petit, soit si l'on a des paquets de petites tailles.

Cette borne de délai est valide pour une topologie arbitraire du réseau, et même si le trafic généré par les autres flots n'est pas régulier du tout. La borne $D_i^{\max}$ n'est donc pas très stricte. Par exemple, des simulations faites par Kurose^[11] d'un réseau avec 4 routeurs, des liens T1 (1.5 Mb/s), et 48 connexions à 32 kb/s, montrent qu'avec une discipline FIFO, plus de 99.9% des paquets atteignent leurs destinations en moins de 15 ms. Mais la borne $D_i^{\max}$ obtenue par Parekh et Gallager^[16] est de 65 ms. Cette borne est une borne garantie, au sens que aucun des paquets ne mettra plus de 65 ms pour atteindre sa destination si les routeurs utilisent WFQ, mais elle n'est pas stricte au sens où la plupart des paquets subissent des délais bien inférieurs à ce qu'elle indique. D'autre part, l'implémentation de l'algorithme WFQ est moins simple que celle de FIFO. Cependant, le résultat de Parekh montre que l'algorithme WFQ permet d'offrir des garanties de bande passante (à savoir la fraction ϕ_i de la bande passante totale au flot i) et de délai (à savoir la valeur $D_i^{\max}$ ci-dessus). Cet algorithme est donc potentiellement très intéressant, et il sera très vraisemblablement implémenté dans l'Internet dans un futur proche. C'est également l'algorithme qui est retenu dans le cadre de notre projet comme discipline d'ordonnancement dans les nœuds du réseau cœur.

1.5 Le contrôle d'admission dans le système CDMA du réseau radio

La capacité de la technique d'accès multiple CDMA est uniquement limitée par le montant d'interférences total généré par les usagers mobiles connectés. Ainsi, CDMA permet d'utiliser les effets du multiplexage sans avoir à prendre en compte la complexité de l'allocation et la réallocation de fréquences comme dans les systèmes FDMA et TDMA. Cependant, dans un système CDMA, il existe une corrélation entre la capacité du système et la qualité de la communication. En effet, le nombre d'usagers occupant une station de

base simultanément doit être limité afin que la qualité de la communication pour chaque usager puisse être maintenue à un niveau correct. Le contrôle d'admission sur l'interface radio a donc un rôle très important dans un systèmes CDMA puisqu'il va contrôler le nombre d'utilisateurs sur les liens radio.

Les exigences de qualité de service sur l'interface radio sont essentiellement la bande passante, le taux d'erreur et le délai. Sur la base de ces exigences qu'il faut satisfaire, des algorithmes basés sur le niveau d'interférence ont été développés dans la littérature. Il s'agit donc d'évaluer le montant d'interférences dans une cellule afin de savoir si l'admission d'un nouvel usager dans cette cellule est possible tout en conservant une bonne qualité de communication pour les utilisateurs en cours.

D'autre part, des procédures de contrôle d'admission ont été développées dans la littérature en considérant que la mobilité sera un critère majeur de qualité de service dans les futurs réseaux radio. La métrique de QoS principale est alors la probabilité pour un utilisateur mobile en situation de *handover* de perdre la communication à cause du changement de station de base le prenant en charge. C'est dans ce contexte qu'ont été développées les techniques de *shadow cluster*^[21] et de partition des canaux^[22].

Dans le cadre de notre étude, la mobilité ne sera pas prise en compte dans le contrôle d'admission sur le réseau radio. Nous nous intéressons donc aux contrôles d'admission basés sur les interférences, c'est-à-dire sur des équations radio permettant d'obtenir une borne sur la capacité du système en termes de puissance. Les utilisateurs sont ainsi admis jusqu'à ce que le montant d'interférence total dépasse un certain seuil. L'estimation de la capacité radio s'appuie essentiellement sur la notion de rapport signal à bruit (SIR). En effet, le SIR du signal en réception détermine la qualité de la communication. Si le signal a un SIR trop faible, alors il sera mal décodé à la réception et sera dégradé. Viterbi *et al.*^[23] puis Liu *et al.*^[24] ont été les premiers à proposer un algorithme basé sur l'estimation du SIR. Ils étudient la capacité radio dans le cas d'un trafic monoservice (de type voix). L'étude est faite sur le lien montant uniquement puisqu'il s'agit du lien le plus limitant en termes de capacité. En effet avec du trafic monoservice symétrique, le lien montant est plus rapidement saturé que le lien descendant à cause de la non orthogonalité des codes dans le sens montant. Plus tard, Landolsi *et al.*^[25] proposent de nouveaux résultats sur

l'estimation de la capacité dans le sens montant.

Plus récemment, avec l'apparition des applications multimédias et des réseaux 3G, de nouvelles équations ont été développées. Elles ont été établies dans un contexte multiservices, et l'étude est souvent faite dans le sens descendant. En effet, les nouvelles applications multimédias seront fortement asymétriques et le lien descendant deviendra le lien limitant. Ainsi Sipila *et al.*^[26] ainsi que Hiltunen *et al.*^[27] proposent une estimation de la puissance totale envoyée par la station de base d'une cellule en s'appuyant sur ce qui a été fait pour le sens montant. De même Cordier *et al.*^[28] donnent une estimation de la puissance totale envoyée et s'appuient sur ce résultat pour donner les limites d'une station de base en termes de capacité et de couverture. Enfin, Akhtar *et al.*^[29] proposent un algorithme de contrôle d'admission se basant sur ces équations en multiservices et prenant en compte la différenciation des services. De la même façon, Jeon *et al.*^[30] proposent un schéma d'admission basé sur le SIR requis par chaque type d'application, prenant en compte à la fois le sens montant et le sens descendant. Holma et Toskala^[4] proposent une évaluation de la capacité radio pour le sens montant et pour le sens descendant dans un système multiservice.

1.6 Le contrôle d'admission dans le réseau cœur

Le but du contrôle d'admission est de limiter le nombre de flux admis dans un réseau afin de satisfaire les demandes de qualité de service des utilisateurs ou applications, tout en partageant les ressources du réseau (bande passante, mémoire tampon) de façon efficace. Le contrôle d'admission est étroitement lié aux autres aspects du réseau tel que les disciplines d'ordonnancement, la caractérisation du trafic et la spécification de la qualité de service.

On peut distinguer deux classes générales de garanties de qualité de service. Les garanties de type déterministes pour le service garanti et les garanties de type statistiques pour le service statistique.

1.6.1 Contrôle d'admission pour le service garanti

Le service garanti^[31] se base sur une caractérisation déterministe du trafic et de la qualité de service désirée. Les garanties de type déterministes se basent sur une description du “pire” comportement possible du trafic. Une métrique de qualité de service sera par exemple: le délai de transit à travers le réseau sera inférieur à 150 ms pour tous les paquets. Les descripteurs de trafic associés à ce service sont simples et la conformité d'un trafic observé avec sa description est facile à vérifier. Cependant l'approche “pire cas” entraîne une surestimation de la bande passante nécessaire et par là-même une sous-utilisation des ressources. Le descripteur de trafic le plus couramment utilisé est le *token bucket*.

La discipline d'ordonancement GPS permet de concevoir des algorithmes de contrôle d'admission pour le service garanti dans le cadre théorique énoncé par Parekh et Gallager^[16]. En effet, en considérant un filtre en “sceau percé” de paramètres σ_i et ρ_i pour la session i , il est possible de déterminer une borne $\hat{D}_i$ sur le temps d'attente maximal des paquets dans les files d'attente d'un réseau de serveurs GPS de topologie arbitraire ainsi qu'une borne $\hat{Q}_i$ sur la longueur maximale des files d'attente allouées à la session i :

$$\hat{Q}_i \leq \sigma_i,$$

$$\hat{D}_i \leq \frac{\sigma_i}{g_i}.$$

Ces bornes pires cas sont simples à calculer et les méthodes de contrôle d'admission se fondent sur ce couplage entre la bande passante allouée g_i et la borne de délai $\hat{D}_i$ pour garantir une limite sur le délai. En effet, la garantie stricte d'une borne sur le délai pour la session i est obtenue en attribuant des taux de services adéquats en chacun des nœuds empruntés par la session. Plus le délai doit être faible, plus la bande passante allouée doit être grande.

Zhang^[32] a étudié plusieurs techniques d'ordonnancement, notamment les approximations de GPS, et a établi pour chacune d'entre elle des bornes pire cas sur le délai et la gigue.

1.6.2 Contrôle d'admission pour le service statistique

Le service statistique se base sur une caractérisation stochastique du trafic et de la qualité de service désirée. Les garanties de type statistique se basent sur une description stochastique du trafic, donc une description du comportement “moyen” du trafic. Les métriques de qualité de service sont définies comme étant la probabilité de dépassement de délai et la probabilité de perte de paquets, par exemple la probabilité que le délai dépasse 150 ms est inférieure à 1%. Il est plus difficile dans ce cas que dans le cas déterministe de vérifier et de faire respecter la conformité du trafic observé avec sa description. Cependant l'approche statistique a l'avantage d'utiliser les ressources plus efficacement. Tous les modèles de trafic se basant sur une bande passante effective sont de ce type.

Les contrôles d'admission pour ce type de service considèrent que chaque flot demande une bande passante comprise entre la valeur moyenne et la valeur crête requises par la session. Cette bande passante est nommée bande passante effective (*effective bandwidth*) et est déterminée grâce aux résultats récents obtenus dans le domaine des grandes déviations (*large deviation theory*). Ainsi la détermination de cette bande passante, qui requiert une connaissance des propriétés statistiques des flots, est complexe. Une caractérisation précise des flots est aussi nécessaire pour réaliser le contrôle d'admission.

Zhang *et al.*^[33] ont proposés différents mécanismes de contrôle d'admission avec de multiples garanties statistiques de QoS pouvant se construire avec la discipline d'ordonnancement GPS, en se basant sur l'analyse probabiliste de la discipline GPS et sur la théorie de la bande passante effective.

1.6.3 Contrôle d'admission pour le service prédictif

Dans l'approche traditionnelle du contrôle d'admission, comme décrite dans les sections 1.6.1 et 1.6.2, l'utilisateur fournit les descripteurs de son trafic, puis à partir de ces descripteurs à priori du trafic, le contrôle d'admission admet ou refuse une session en se basant sur le calcul du comportement pire cas pour le service garanti, et sur le calcul de la bande passante effective pour le service statistique. L'utilisateur est donc responsable de la caractérisation de son trafic. Or ceci est difficile en général car il n'est pas toujours

possible de prédire ce que sera le comportement du flot.

Le contrôle d'admission à base de mesures (MBCAC), contrairement à l'approche traditionnelle, estime lui-même les paramètres du modèle de trafic, et ne se base plus sur des descripteurs de trafic à priori. De ce fait, le MBCAC permet une meilleure utilisation des ressources que pour le service garanti, toutefois la garantie sur le délai sera moins fiable. D'autre part, le MBCAC est intrinsèquement une procédure statistique puisqu'il se base sur des mesures, et il ne peut donc s'appliquer uniquement dans le cadre de garanties statistiques et de modèles de trafic stochastique.

Le contrôle d'admission à base de mesures est donc utilisé pour du service prédictif. On parle de service prédictif car l'on considère que les mesures du passé sont un bon indicateur du comportement futur du réseau. Or ceci peut s'avérer dangereux car si cette prédiction du comportement futur du trafic n'est pas bonne, le MBCAC risque soit de sous-utiliser les ressources du réseau, soit d'admettre trop de flots et dégrader ainsi les qualités de service demandées par les sessions en cours. Ainsi ce type de contrôle d'admission est bien approprié pour le service à contrôle de charge^[34]. Ce service ne fait pas de garanties absolues sur le service demandé, il considère une borne sur le délai nominale puis autorise des variations autour de cette valeur. Le but est d'obtenir une qualité de service semblable à celle que la session aurait eu dans des conditions où le réseau est peu chargé.

1.6.3.1 Les différentes techniques de mesure

Un certain nombre d'études ont conduit à l'exploration de différentes techniques de mesure du chargement du réseau. La première technique de mesure est le mécanisme de fenêtre temporelle (*time-window*) utilisé par Jamin *et al.*^[35]. Ils calculent un débit moyen sur chaque période d'échantillonnage de taille p . A la fin de la fenêtre de mesure de taille T , ils prennent la plus grande valeur moyenne calculée sur cette fenêtre comme l'estimation du chargement pour la durée T de la fenêtre suivante. Lorsqu'un nouveau flot est admis dans le réseau, cette estimation est immédiatement augmentée de la valeur du débit déclaré par ce flot. En tout temps, si la nouvelle valeur moyenne du débit calculée sur une période p est supérieure à la valeur de l'estimation du chargement, la valeur de

l'estimation est immédiatement augmentée de la valeur de cette nouvelle moyenne.

Le deuxième mécanisme de mesure est le mécanisme *point samples*. Il prend la valeur moyenne du chargement sur chaque période p' comme estimation du chargement du réseau. Il s'agit en fait du mécanisme de fenêtre temporelle énoncé précédemment avec T/p qui vaut 1.

Le troisième et dernier mécanisme est le mécanisme *exponential averaging*. Une valeur moyenne du chargement $\hat{\nu}^p$ est calculée sur chaque période d'échantillonnage p . Puis l'estimation du chargement est la valeur $\hat{\nu}'$ suivante: $\hat{\nu}' = (1 - w) * \hat{\nu} + w * \hat{\nu}^p$, calculée à la fin de chaque période p . w est un nombre compris entre 0 et 1, et $\hat{\nu}$ est la valeur de l'estimation du chargement calculée à la période précédente.

1.6.3.2 Les différents algorithmes MBCAC

Chaque algorithme MBCAC proposé dans la littérature se place dans le contexte où les ressources pour chaque flot sont réservées à l'aide d'un protocole de réservation, par exemple le protocole RSVP. Chaque flot est décrit par les paramètres de son *token bucket*, son taux de remplissage, sa profondeur et la valeur de débit crête. La plupart des algorithmes proposés font des mesures sur le trafic agrégé. La qualité de service est exprimée en perte de paquets, et les études consistent à établir le taux d'utilisation de la bande passante pour une valeur de perte de paquets donnée. Cette courbe peut être obtenue en jouant sur les paramètres de l'algorithme de mesure qui influencent grandement la performance du MBCAC. Des comparaisons des algorithmes les plus importants ont été faites dans la littérature^{[36][37][38]} et sont récapitulées dans le tableau 1.5, où $\hat{\nu}$ représente le débit agrégé mesuré des flots dans le système et r la bande passante disponible sur le nœud où se fait l'admission.

La plupart des algorithmes proposés sont assez similaires au niveau de leur équation d'admission qui est de la forme: $\hat{\nu} < f(.)r - g(.)$, où $f(.)$ et $g(.)$ sont des fonctions qui dépendent du débit réservé par la source et du nombre de flots admis. Il a été établi que la performance de ces algorithmes n'est pas sensible à la forme des équations d'admission, mais plutôt aux changements de valeurs des paramètres de la technique de mesure. Il existe deux objectifs de performance pour un MBCAC: le premier est d'atteindre le plus

TAB. 1.5 – Comparaison des différents MBCAC

		Condition d'admission	Technique de mesure
<i>Measured Sum</i> ^[35]	Ad-hoc	$\hat{v} + \rho < vr$ v : facteur d'utilisation	<i>Time Window</i> Longueur de fenêtre fixe ^[35] Longueur de fenêtre variable ^{[39][40]}
<i>Aggregate Traffic Envelope</i> ^[41]	Ad-hoc	Utilise des mesures de moyenne et de variance des enveloppes de trafic	<i>Time Window</i>
Bande passante équivalente:			
<i>Hoeffding Bounds</i> ^[42]	Théorie mathématique	$\hat{C}_H(\hat{v}, n) + p \leq r$ $\hat{C}_H$: bande passante effective mesurée n : nombre de flots	<i>Exponential averaging</i>
<i>Measure CAC</i>	Théorie mathématique	$\hat{C}_H + p \leq r$ $\hat{C}_H$ calculée à partir du taux de perte demandé et de la fonction génératrice des arrivées des flux	<i>Exponential averaging</i>
<i>Tangent at Peak</i>	Théorie mathématique	$np(1 - e^{-sp}) + e^{-sp}\hat{v} \leq r$ n : nombre de flots s : paramètre spacial des bornes de <i>Chernoff</i>	<i>Point sample</i>
<i>Tangent at Origin</i>	Théorie mathématique	$e^{sp}\hat{v} \leq r$ s : paramètre spacial des bornes de <i>Chernoff</i>	<i>Point sample</i>

haut taux d'occupation de la bande passante étant donné un certain taux de perte de paquets, et le deuxième est d'adhérer au mieux au service demandé par la source. En ce qui concerne le premier objectif, il a été montré que les différents algorithmes produisent quasiment tous les mêmes performances, même s'ils sont issus de théories différentes. Pour le second objectif, aucun des algorithmes n'atteint véritablement les critères de performance en termes de qualité du service demandé.

Grossglauser *et al.*^[43] ont à ce sujet travaillé à l'élaboration de directives pour le développement d'un MBCAC robuste, c'est-à-dire capable de donner les mêmes garanties de QoS qu'un contrôle qui se base sur la connaissance exacte des spécifications de trafic. Ceci leur est revenu à comprendre l'impact de l'incertitude implicite dans chaque procédure d'estimation sur la performance d'un MBCAC. Ils montrent en particulier qu'un contrôle d'admission qui fait une hypothèse de certitude équivalente, et donc suppose que les paramètres mesurés sont les vrais paramètres, mène à une dégradation de la qualité de service. Ils cherchent donc à compenser cette dégradation, en comprenant notamment l'impact sur la qualité de service i) de l'incertitude due à l'estimation des paramètres, ii) des caractéristiques du trafic (en particulier des corrélations temporelles), et iii) de la dynamique du système causée par les arrivées et les départs de flux. Ils montrent ainsi

que si la taille de la fenêtre de mesure T est choisie du même ordre de grandeur que l'échelle temporelle critique $\tilde{T}_h = \frac{T_h}{\sqrt{n}}$, où T_h est la durée moyenne d'un flux et $n = \frac{r}{\mu}$ avec r la capacité du serveur et μ le débit moyen d'un flux, alors on obtient un contrôle d'admission robuste.

1.7 Conclusion

La qualité de service et le trafic multiservices sont les concepts clés des réseaux de troisième génération. Afin de gérer la qualité de service, nous retenons l'importance du contrôle d'admission en entrée du réseau cœur qui est lié à la politique d'ordonnancement des paquets en chacun des nœuds du réseau, la discipline WFQ restant la plus prometteuse. De plus, en entrée du réseau radio, le contrôle d'admission permet également la gestion du trafic et de la qualité de service. Il s'appuie sur les propriétés de la technique de multiplexage CDMA, prend en compte la nature multiservice du trafic ainsi que les différents mécanismes radios tels que le *soft handover* et le contrôle de puissance.

Dans ce mémoire, nous proposons un contrôle d'admission pour du service garanti et intégrant des mesures afin d'optimiser la taux d'utilisation de la bande passante. Il s'appuie sur les propriétés de la discipline WFQ développées dans ce chapitre pour le réseau cœur, et sur les équations de capacité multiservices pour le réseau d'accès radio. Le développement de ces équations est l'objet du chapitre suivant.

CHAPITRE 2

EVALUATION DE LA CAPACITÉ RADIO CDMA2000

PREFACE

Ce chapitre présente les formules de capacité radio élaborées par un ensemble de personnes durant le déroulement de la maîtrise. Elles sont issues d'un premier travail de Cyrille Klam et Monsieur Roland Malhamé qui ont développé des formules multiservices en s'inspirant du mémoire de Steeve Tsang-Kwong-U^[44] ayant travaillé dans un contexte monoservice. Par la suite, durant cette maîtrise, ces équations se sont complexifiées et ont évolué, les hypothèses en ont été mieux cernées par le fruit d'un travail de collaboration entre différents acteurs: moi-même, Monsieur Roland Malhamé, Madame Brigitte Jau-mard, et Monsieur Francis Lupien d'Ericsson. Au travers de nombreuses discussions et rencontres où j'ai pris activement part, les équations ont pris la forme théorique présentée dans ce mémoire. J'ai ensuite analysé ces équations en détail, en ait cerné certaines limitations importantes, et en ait fait une validation numérique et une analyse pratique. Cela m'a permis ensuite de pouvoir utiliser ces formules dans le contexte du contrôle d'admission développé dans ce mémoire.

Pour décider de l'admission des sessions dans le réseau d'accès radio, nous utilisons des modèles originaux d'évaluation de la capacité CDMA dans un contexte multiservices, développés dans le cadre de ce projet par Jaumard *et al.*^[45]. En s'appuyant sur les estimations de la capacité radio développés dans la littérature et référencées à la section 1.5, nous proposons un modèle multiservices pour le sens montant et pour le sens descendant. Le but est ainsi de mettre en évidence les conditions dans lesquelles le système est en état de congestion. Le test de ces conditions permet alors de savoir si une nouvelle session se présentant à l'entrée du réseau d'accès radio peut ou non être admise. La congestion a lieu lorsque le montant d'interférences dans le système CDMA dépasse un certain seuil. Notons que dans cette étude la mobilité des usagers n'est pas prise en compte dans le contrôle d'admission. Le contexte de l'environnement radio CDMA2000 dans lequel nous travaillons est décrit dans la section 1.3.

2.1 Estimations de la capacité CDMA dans la littérature

La table 2.1 rappelle les principales équations de capacité CDMA de la littérature. Notons que peu de références ont traité le *soft handover* dans leurs équations. Ce tableau permet, après la lecture des sections qui suivent, de situer les équations développées dans notre étude par rapport aux équations déjà existantes. L'ensemble des références sur la capacité radio est donné à la section 1.5. Dans le tableau, toutes les notations sont celles présentées dans les sections suivantes. Il est ajouté ici les notations I_0 qui représente la densité spectrale d'interférences maximum tolérable pour le sens montant, ϕ_i qui exprime le fait que chaque station de base i envoie une puissance $\phi_i P_{BS}$ pour le sens descendant, et N_{BS} qui représente le nombre de stations de bases qui interfèrent dans le sens descendant.

2.2 Modèle d'évaluation de la capacité sur le lien montant

Nous proposons ici un modèle mathématique d'estimation de la capacité d'une cellule dans le réseau d'accès radio pour le sens montant. Il s'agit d'établir l'expression mathématique de la puissance totale reçue à la station de base de la cellule en question.

TAB. 2.1 – Les équations de capacité radio dans la littérature

monoservice - sens montant	
Viterbi <i>et al.</i> ^[23]	$\sum_{s=1}^N \nu_s E_{b_s} R_s + \sum_i^{\text{other cells}} \sum_{s=1}^N \nu_{s(i)} E_{b_{s(i)}} R_s + N_0 W \leq I_0 W$
multiservices - sens montant	
Holma et Toskala ^[4]	$\eta_{UL} = (1 + \lambda) \sum_{s=1}^N \frac{1}{1 + \frac{W}{(E_{b_s}/N_0)R_s \nu_s}} < 1$
multiservices - sens descendant	
Holma et Toskala ^[4]	$\eta_{DL} = \sum_{s=1}^N \nu_s \frac{(E_{b_s}/N_0)}{W/R_s} [w_s + \lambda_s] < 1$
Sipila <i>et al.</i> ^[26]	$\eta_{DL} = \sum_{s=1}^N \frac{(E_{b_s}/N_0) \nu_s R_s}{W} \left(w_s + \sum_{i=1}^{N_{BS}} \frac{a_{m_s,0}}{a_{m_s,i}} \right) < 1$
Ahktar <i>et al.</i> ^[29]	$\sum_{s=1}^N P_s = \sum_{s=1}^N \frac{\left(\frac{\nu_s R_s}{W}\right) \gamma_s P_{BS}}{1 + w \left(\frac{\nu_s R_s}{W}\right) \gamma_s} \left[w \phi_0 + \frac{1}{a_{m_s,0}} \sum_{i=1}^{N_{BS}} \phi_i a_{m_s,i} \right] \leq P_{BS}$

2.2.1 Hypothèses

Nous présentons ici les différentes hypothèses que nous faisons pour élaborer l'équation de capacité dans le sens montant.

- Les interférences intra-cellulaires et inter-cellulaires sont prises en compte dans le modèle. Cependant, les interférences inter-cellulaires sont considérées comme étant un pourcentage de la puissance totale reçue à la station de base.
- Le contrôle de puissance est supposé être parfait.
- Les puissances liées au signal pilote et au signal de contrôle sur le lien montant ne sont pas prises en compte.
- Les cellules ne sont pas sectorisées, c'est-à-dire que l'antenne au niveau de la station de base est omni-directionnelle. Chaque cellule ne possède donc qu'un seul secteur. On suppose également qu'il n'y a qu'une antenne par station de base.
- Nous supposons que le mobile n'est pas limité par sa puissance d'émission.

2.2.2 Définition des paramètres

L'abréviation UL désigne le sens montant.

s	session active initiée par un mobile dans la cellule considérée.
S	ensemble des sessions actives dans la cellule considérée.
RAB_s^{UL}	RAB de la session s , dépendant du type d'application considéré et donc caractérisé par sa bande passante R_s^{UL} et son taux d'erreur par trame FER_s^{UL} .
P_s^{UL}	puissance reçue à la station de base de la part d'une session $s \in S$.
ν_s^{UL}	facteur d'activité associé à la session s .
P_T^{UL}	puissance totale reçue à la station de base de la part de toutes les sessions actives dans la cellule.
λ	fraction de la puissance totale P_T^{UL} représentant les interférences inter-cellulaires. Holma et Toskala ^[4] proposent de prendre 0.55 pour la valeur de λ .
N_0	densité spectrale de puissance du bruit thermique.
W	bande passante de CDMA2000.
$\left(\frac{E_b}{N_t}\right)_s^{UL}$	rapport de l'énergie par bit du signal reçu à la station de base de la part de la session s sur la densité spectrale de puissance du bruit reçu (bruit thermique + interférences).
γ_s^{UL}	valeur cible de $\left(\frac{E_b}{N_t}\right)_s^{UL}$ que doit avoir le signal reçu à la station de base de la part de la session s pour que le décodage du signal soit possible.

2.2.3 Mise en équations du modèle sans *soft handover*

Dans le sens montant, le signal émis par la session s d'un mobile va pouvoir être récupéré à la station de base si le rapport $\left(\frac{E_b}{N_t}\right)_s^{UL}$ du signal reçu à la station de base de la part de cette session est suffisamment grand. Plus précisément, étant donné une

session s caractérisée par son RAB RAB_s^{UL} , il existe des abaques permettant de connaître la valeur de $\left(\frac{E_b}{N_t}\right)_s^{UL}$ nécessaire pour que le signal soit récupérable à la station de base. On a noté γ_s^{UL} cette valeur. L'annexe I détaille la lecture de ces abaques. Ainsi le mobile initiant la session s doit émettre suffisamment de puissance afin que la puissance reçue à la station de base P_s^{UL} soit assez grande pour assurer un rapport $\left(\frac{E_b}{N_t}\right)_s^{UL}$ égal à γ_s^{UL} . Nous supposons que pour chaque session s active le mobile est capable d'émettre assez de puissance pour atteindre la valeur cible γ_s^{UL} à la station de base.

Le rapport signal à bruit noté SIR_s^{UL} est relié au rapport énergie par bit du signal sur densité spectrale du bruit de la façon suivante:

$$\left(\frac{E_b}{N_t}\right)_s^{UL} = SIR_s^{UL} \frac{W}{R_s^{UL}}$$

avec

- $SIR_s^{UL} = \frac{\text{puissance du signal de la session } s \text{ reçu à la station de base}}{\text{puissance de tous les autres signaux reçus à la station de base}},$
- $\frac{W}{R_s^{UL}}$ est appelé facteur d'étalement.

Ainsi, dans un contexte monoservice, toutes les sessions ont le même RAB attribué, et le SIR_s^{UL} ou le $\left(\frac{E_b}{N_t}\right)_s^{UL}$ requis est le même pour toutes les sessions. En revanche dans un contexte multiservices, le $\left(\frac{E_b}{N_t}\right)_s^{UL}$ requis et lu sur les abaques diffère suivant le type d'application.

Maintenant que $\left(\frac{E_b}{N_t}\right)_s^{UL}$ est clairement défini, nous pouvons en donner l'expression mathématique:

$$\left(\frac{E_b}{N_t}\right)_s^{UL} = \frac{\frac{P_s^{UL}}{R_s^{UL}}}{\frac{1}{W} ((1 + \lambda) P_T^{UL} - \nu_s^{UL} P_s^{UL}) + N_0} \quad (2.1)$$

avec

- $P_T^{UL} - \nu_s^{UL} P_s^{UL}$ représente les interférences intra-cellulaires subies par la session s ,
- λP_T^{UL} représente les interférences inter-cellulaires.

On remarque qu'au numérateur on considère la puissance du signal de la session s lorsque cette dernière est active, on ne tient pas compte du facteur d'activité car pour le décodage du signal à la station de base il faut considérer la puissance reçue de la part de

la session s dans une phase d'activité. Par contre, au dénominateur, nous exprimons les interférences moyennes subies par la session s , nous pouvons donc considérer la puissance moyenne du signal reçu prenant en compte les périodes d'inactivité de la session s .

On peut donc maintenant extraire de l'équation (2.1) l'expression de P_s^{UL} sachant que l'on veut un rapport énergie par bit du signal sur densité spectrale du bruit égale à γ_s^{UL} pour la session s :

$$P_s^{\text{UL}} = \gamma_s^{\text{UL}} \frac{(1 + \lambda) \frac{P_T^{\text{UL}}}{W} + N_0}{\frac{1}{R_s^{\text{UL}}} + \frac{\nu_s^{\text{UL}} \gamma_s^{\text{UL}}}{W}}. \quad (2.2)$$

De plus nous savons que:

$$P_T^{\text{UL}} = \sum_{s \in S} \nu_s^{\text{UL}} P_s^{\text{UL}}. \quad (2.3)$$

Ainsi à partir des équations (2.2) et (2.3) on obtient l'expression de P_T^{UL} , soit la valeur que prend P_T^{UL} si toutes les sessions s actives sont décodables à la station de base:

$$P_T^{\text{UL}} = \frac{\sum_{s \in S} \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{\frac{1}{R_s^{\text{UL}}} + \frac{\nu_s^{\text{UL}} \gamma_s^{\text{UL}}}{W}}}{1 - \frac{1 + \lambda}{W} \sum_{s \in S} \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{\frac{1}{R_s^{\text{UL}}} + \frac{\gamma_s^{\text{UL}}}{W}}} N_0. \quad (2.4)$$

Observons maintenant le dénominateur de l'équation (2.4). On s'aperçoit que si ce dénominateur devient proche de 0 alors le niveau de bruit (*noise rise*) devient proche de l'infini et le système atteint alors ce que l'on appelle sa *pole capacity*. En effet, le *noise rise* est le rapport de la puissance totale reçue à la station de base, bruit thermique y compris, sur la puissance du bruit thermique. Donc l'expression du *noise rise* est la suivante:

$$\text{noise_rise} = \frac{P_T^{\text{UL}} + N_0 W}{N_0 W} = \frac{1}{1 - \eta_{\text{UL}}} \quad (2.5)$$

avec

$$\eta_{\text{UL}} = \frac{1 + \lambda}{W} \sum_{s \in S} \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{\frac{1}{R_s^{\text{UL}}} + \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{W}}. \quad (2.6)$$

η_{UL} est ce qu'on appelle le facteur de charge *load factor*. Ainsi si η_{UL} devient proche de 1, le système atteint sa *pole capacity* et devient instable. La puissance totale reçue à la station de base devient en effet très supérieure à la puissance du bruit thermique, et le décodage à la station de base n'est plus possible. Il faut donc faire en sorte que:

$$\eta_{UL} = \frac{1 + \lambda}{W} \sum_{s \in S} \frac{\gamma_s^{UL} \nu_s^{UL}}{\frac{1}{R_s^{UL}} + \frac{\gamma_s^{UL} \nu_s^{UL}}{W}} < LF_{UL} \quad (2.7)$$

avec $0 < LF_{UL} < 1$. Holma et Toskala^[4] suggèrent d'avoir un *noise rise* d'environ 4dB, cela correspondant à un *load factor* de 60%.

Ainsi nous appellerons capacité (sans *soft handover*) d'une session s demandant l'entrée dans le réseau radio dans le sens montant la valeur : $\frac{1+\lambda}{W} \frac{\gamma_s^{UL} \nu_s^{UL}}{\frac{1}{R_s^{UL}} + \frac{\gamma_s^{UL} \nu_s^{UL}}{W}}$. Puis si la somme des capacités des sessions actives et de la session s faisant la requête d'admission est inférieure à LF_{UL} alors la session s pourra être admise.

2.2.4 Equations avec *soft handover*

Les mobiles en *soft handover* bénéficient du service d'au moins deux antennes, comme décrit à la section 1.3.3. Pour les calculs qui suivent on considère seulement deux antennes actives pour un mobile en *soft handover*. Le mobile envoie son signal aux deux antennes, chaque antenne décode le signal reçu et l'envoie au BSC (les deux antennes étant reliées au même BSC). Le BSC récupère donc deux signaux et reconstitue alors le signal envoyé par le mobile en sélectionnant à chaque fois la trame avec le taux d'erreur le plus faible. Ainsi le lien radio devient beaucoup plus fiable si le mobile est servi par deux antennes que s'il est servi par une seule antenne car la probabilité de détecter correctement le signal est augmentée. En conséquence, le signal envoyé par un mobile en *soft handover* a un $\frac{E_b}{N_t}$ requis moins important que s'il n'était pas en *soft handover* et il peut ainsi envoyer moins de puissance. La situation géométrique qui maximise le gain apporté par le service de deux antennes est un mobile placé exactement au milieu entre les deux antennes. Lorsqu'un mobile en *soft handover* n'est pas à égale distance des stations de base qui le prennent en charge, ce gain en macro-diversité diminue. Ainsi, Holma et Toskala^[4] estiment que l'on a un gain de 2.7dB à 4dB sur la valeur de $\frac{E_b}{N_t}$ requise pour un usager à

pied (3km/h), dépendant du facteur d'atténuation entre lui et ses stations de base.

Nous définissons SHG_{UL} le facteur de réduction de la valeur de $\frac{E_b}{N_t}$ pour un mobile en *soft handover*. Notons $p_{\text{SH}}^{\text{UL}}$ la proportion de sessions actives initiées par des mobiles en *soft handover* dans une cellule. Il est alors important de préciser que la valeur de $p_{\text{SH}}^{\text{UL}}$ est une constante que nous fixons, par exemple à 20%, ce qui représente une limitation du modèle. En effet, en réalité plus il y a des mobiles en *soft handover* dans une cellule, plus la cellule voisine, qui a les mêmes caractéristiques de trafic, a aussi un nombre important de mobiles en *soft handover*, ce qui contribue à augmenter les interférences inter-cellulaires dans la cellule que l'on considère. Or dans notre étude, la proportion de sessions en *soft handover* est indépendante du montant d'interférences inter-cellulaires qui est lui-même une fraction fixe de la puissance totale. Ainsi dans nos équations, plus nous considérons de mobiles en *soft handover*, plus ce modèle va accepter des mobiles dans le sens montant grâce au gain en macrodiversité, alors qu'en réalité plus il y a de mobiles dans cette position, plus les interférences inter-cellulaires sont importantes, ce qui a l'effet opposé du gain en macrodiversité. Dans notre modélisation il convient d'utiliser une valeur de $p_{\text{SH}}^{\text{UL}}$ raisonnable autour de 20% afin que la formule donne des résultats satisfaisants.

Pour chaque session s nous allons utiliser une valeur moyenne du $\left(\frac{E_b}{N_t}\right)_s^{\text{UL}}$ requis définie comme suit:

$$\gamma_{\text{moy};s}^{\text{UL}} = p_{\text{SH}}^{\text{UL}}(\text{SHG}_{\text{UL}}\gamma_s^{\text{UL}}) + (1 - p_{\text{SH}}^{\text{UL}})\gamma_s^{\text{UL}},$$

γ_s^{UL} étant la valeur de $\left(\frac{E_b}{N_t}\right)_s^{\text{UL}}$ requise pour la session s sans le *soft handover* (donc lue sur les abaques).

On a alors la nouvelle contrainte de capacité:

$$\frac{1 + \lambda}{W} \sum_{s \in S} \frac{\gamma_{\text{moy};s}^{\text{UL}} \nu_s^{\text{UL}}}{\frac{1}{R_s^{\text{UL}}} + \frac{\gamma_{\text{moy};s}^{\text{UL}} \nu_s^{\text{UL}}}{W}} < LF_{\text{UL}},$$

soit

$$\frac{1 + \lambda}{W} \sum_{s \in S} \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{\frac{1}{(1 - p_{\text{SH}}^{\text{UL}}(1 - \text{SHG}_{\text{UL}}))R_s^{\text{UL}}} + \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{W}} < LF_{\text{UL}}. \quad (2.8)$$

Ainsi lorsque le *soft handover* est pris en compte, la capacité d'une session s devient:

$$\frac{1+\lambda}{W} \frac{\gamma_s^{UL} \nu_s^{UL}}{\left(\frac{1}{(1-P_{SH}^{UL}(1-SHG_{UL}))} R_s^{UL} + \frac{\gamma_s^{UL} \nu_s^{UL}}{W} \right)}.$$

2.3 Modèle d'évaluation de la capacité sur le lien descendant

Nous proposons ici un modèle mathématique d'estimation de la capacité d'une cellule dans le réseau d'accès radio pour le sens descendant. Il s'agit d'établir l'expression mathématique de la puissance totale envoyée par la station de base de la cellule en question vers ses mobiles.

2.3.1 Hypothèses

Nous présentons ici les hypothèses que nous faisons pour développer l'équation de capacité dans le sens descendant.

- On considère un plan infini de cellules où les stations de base (une par cellule) sont équidistantes les unes des autres (voir annexe II).
- Les interférences intra-cellulaires et inter-cellulaires sont prises en compte.
- Seul le premier anneau des cellules avoisinantes (6 cellules) est pris en compte dans l'évaluation des interférences inter-cellulaires.
- L'intensité du trafic et la distribution des usagers est supposée être identique dans chaque cellule.
- L'atténuation des signaux se base sur un modèle simple de propagation avec onde directe et onde réfléchie comme décrit dans l'annexe II.
- Le contrôle de puissance est supposé parfait.
- Les puissances liées aux signaux de contrôle et de synchronisation sont prises en compte.
- Les cellules ne sont pas sectorisées. On suppose également qu'il n'y qu'une antenne par station de base.

2.3.2 Définition des paramètres

L'abréviation DL désigne le sens descendant.

s	session active dans la cellule considérée.
S	ensemble des sessions actives dans la cellule considérée.
m_s	mobile récepteur du signal de la session s .
RAB_s^{DL}	RAB de la session s , dépendant du type d'application, et donc caractérisé par sa bande passante R_s^{DL} et son taux d'erreur par trame FER_s^{DL} .
P_s^{DL}	puissance envoyée par la station de base pour la session $s \in S$.
ν_s^{DL}	facteur d'activité associé à la session s .
P_T^{DL}	puissance totale envoyée par la station de base, signaux de contrôle y compris.
N_0	densité spectrale de puissance du bruit thermique.
W	bande passante de CDMA2000.
$\left(\frac{E_b}{N_t}\right)_s^{DL}$	rapport de l'énergie par bit du signal reçu au mobile m_s sur la densité spectrale de puissance du bruit reçu à ce mobile (bruit thermique + interférences).
γ_s^{DL}	valeur cible de $\left(\frac{E_b}{N_t}\right)_s^{DL}$ que doit avoir le signal reçu au mobile m_s pour que le décodage du signal soit possible.
$a_{m_s,0}$	coefficient d'atténuation du signal entre le mobile m_s et la station de base qui le prend en charge.
P_{BS}	puissance maximale disponible à la station de base, que l'on prend égale à 20 W.
P_{cont}	puissance des signaux de contrôle et synchronisation considérée comme étant

une fraction F_{cont} de la puissance P_{BS} . Nous choisissons de prendre

$$F_{cont} = 0,2.$$

- $a_{m_s,i}$ coefficient d'atténuation du signal entre le mobile m_s et la station de base i .
- $\mathcal{B}$ ensemble des 6 stations de base immédiatement voisines de la station de base de la cellule considérée.
- w facteur d'orthogonalité. Les codes orthogonaux sont utilisés dans le sens descendant afin de différencier les usagers, cependant à la réception au niveau des mobiles ces codes ne sont plus parfaitement orthogonaux à cause de la propagation sur des chemins multiples. $w = 0$ correspond à l'orthogonalité parfaite des codes. Holma et Toskala^[4] le prennent typiquement entre 0.1 et 0.4.

2.3.3 Mise en équations du modèle sans *soft handover*

Dans le sens descendant, le signal émis par la session s à la station de base va pouvoir être récupéré au mobile m_s si le rapport $\left(\frac{E_b}{N_t}\right)_s^{DL}$ du signal reçu au mobile est suffisamment grand. Plus précisément, étant donné une session s caractérisée par son RAB RAB_s^{DL} , il existe des abaques (différentes de celles pour le sens montant) permettant de connaître la valeur de $\left(\frac{E_b}{N_t}\right)_s^{DL}$ nécessaire pour que le signal soit récupérable à la réception au mobile. On a noté γ_s^{DL} cette valeur. L'annexe I détaille la lecture de ces abaques. Ainsi la station de base doit émettre suffisamment de puissance P_s^{DL} pour la session s afin que la puissance reçue au mobile soit assez grande pour assurer un rapport $\left(\frac{E_b}{N_t}\right)_s^{DL}$ égal à γ_s^{DL} .

De la même façon que dans le sens montant, on a la relation suivante:

$$\left(\frac{E_b}{N_t}\right)_s^{DL} = SIR_s^{DL} \frac{W}{R_s^{DL}}$$

avec $SIR_s^{DL} = \frac{\text{puissance du signal de la session } s \text{ reçu au mobile}}{\text{puissance de tous les autres signaux reçus au mobile}}.$

$\left(\frac{E_b}{N_t}\right)_s^{\text{DL}}$ étant défini, on peut en donner l'expression mathématique suivante:

$$\left(\frac{E_b}{N_t}\right)_s^{\text{DL}} = \frac{\frac{1}{R_s^{\text{DL}}} \frac{P_s^{\text{DL}}}{a_{m_s,0}}}{\frac{1}{W} \left(w \left(\frac{P_T^{\text{DL}} - \nu_s^{\text{DL}} P_s^{\text{DL}}}{a_{m_s,0}} \right) + \sum_{i \in \mathcal{B}} \frac{P_T^{\text{DL}}}{a_{m_s,i}} \right) + N_0}, \quad (2.9)$$

- $w \left(\frac{P_T^{\text{DL}} - \nu_s^{\text{DL}} P_s^{\text{DL}}}{a_{m_s,0}} \right)$ représente les interférences intra-cellulaires subies par la session s ,
- $\sum_{i \in \mathcal{B}} \frac{P_T^{\text{DL}}}{a_{m_s,i}}$ représente les interférences inter-cellulaires.

On peut donc maintenant extraire de l'équation (2.9) l'expression de P_s^{DL} sachant que l'on veut un rapport énergie par bit du signal sur densité spectrale du bruit égale à γ_s^{DL} pour la session s :

$$P_s^{\text{DL}} = \gamma_s^{\text{DL}} \frac{P_T^{\text{DL}} \left(w + \sum_{i \in \mathcal{B}} \frac{a_{m_s,0}}{a_{m_s,i}} \right) + N_0 W a_{m_s,0}}{\frac{W}{R_s^{\text{DL}}} + w \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}. \quad (2.10)$$

De plus la puissance envoyée par la station de base est:

$$P_T^{\text{DL}} = \sum_{s \in S} \nu_s^{\text{DL}} P_s^{\text{DL}} + P_{\text{cont}}. \quad (2.11)$$

Ainsi à partir des équations (2.10) et (2.11) on obtient l'expression de P_T^{DL} , soit la valeur que prend P_T^{DL} si toutes les sessions s actives sont décodables à leur mobile récepteur:

$$P_T^{\text{DL}} = \frac{P_{\text{cont}} + W N_0 \sum_{s \in S} \frac{\nu_s^{\text{DL}} \gamma_s^{\text{DL}} a_{m_s,0}}{\frac{W}{R_s^{\text{DL}}} + w \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}}{1 - \sum_{s \in S} \frac{\left(w + \sum_{i \in \mathcal{B}} \frac{a_{m_s,0}}{a_{m_s,i}} \right) \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}}. \quad (2.12)$$

De la même manière que dans le sens montant, on peut mettre en évidence le facteur de charge (*load factor*) dans le sens descendant:

$$\eta_{\text{DL}} = \sum_{s \in S} \frac{\left(w + \sum_{i \in \mathcal{B}} \frac{a_{m_s,0}}{a_{m_s,i}} \right) \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}.$$

Lorsque η_{DL} devient proche de 1, le système devient instable, et P_T^{DL} devient infini. En fait, dans le sens descendant, la puissance émise par la station de base P_T^{DL} est limitée rapidement par la puissance d'émission maximale de l'antenne P_{BS} .

La contrainte la plus forte dans le sens descendant est donc:

$$P_T^{DL} < P_{BS}. \quad (2.13)$$

Cependant, dans notre étude nous ne connaissons pas les positions exactes des usagers dans la cellule, nous ne connaissons que la distribution stochastique des usagers dans la cellule: nous supposons que les usagers sont répartis uniformément dans une région circulaire de rayon R autour de la station de base de la cellule (voir l'annexe II). Ainsi les coefficients d'atténuation sont des variables aléatoires dont on connaît les espérances à partir de la fonction de distribution des positions des usagers sur la surface circulaire de rayon R . On notera $E[a_{m_s,0}]$ la valeur moyenne de l'atténuation $a_{m_s,0}$ et $E[\frac{a_{m_s,0}}{a_{m_s,i}}]$ la valeur moyenne du rapport des atténuations $\frac{a_{m_s,0}}{a_{m_s,i}}$ (cette valeur est la même $\forall i \in \mathcal{B}$ car on suppose les usagers mobiles répartis uniformément). Pour simplifier les calculs nous allons supposer que le signal de chaque session s subit une atténuation de la station de base au mobile m_s égale à $E[a_{m_s,0}]$, et que pour chaque session s le rapport des pertes $\frac{a_{m_s,0}}{a_{m_s,i}}$ est égal à la valeur moyenne $E[\frac{a_{m_s,0}}{a_{m_s,i}}]$. Cela suppose donc que tous les mobiles se trouvent sur un rayon moyen. Le calcul des coefficients d'atténuation pour une distribution uniforme des usagers est détaillée dans l'annexe II.

D'après l'équation (2.12) et sous l'approximation que l'on vient de faire on a:

$$P_T^{DL} = \frac{P_{cont} + WN_0 \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL} E[a_{m_s,0}]}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}}{1 - \sum_{s \in S} \frac{\left(w + \sum_{i \in \mathcal{B}} E[\frac{a_{m_s,0}}{a_{m_s,i}}] \right) \nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}}.$$

On a ainsi:

$$P_T^{DL} = \frac{P_{cont} + WN_0 E[a_{m_s,0}] \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}}{1 - \left(w + 6 E[\frac{a_{m_s,0}}{a_{m_s,i}}] \right) \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}}. \quad (2.14)$$

On exprime maintenant la condition 2.13 en introduisant le facteur LF_{DL} ($0 < LF_{DL} < 1$) afin de prendre en compte l'erreur que l'on fait en approximant les termes d'atténua-tions par leurs valeurs moyennes, il est raisonnable de prendre LF_{DL} égal à 0.6:

$$\frac{P_{cont} + WN_0 E[a_{m_s,0}] \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}}{1 - \left(w + 6 \cdot E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] \right) \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}} < LF_{DL} \cdot P_{BS},$$

ce qui nous donne:

$$\alpha \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}} < DL_{threshold} \quad (2.15)$$

avec

$$DL_{threshold} = P_{BS} (LF_{DL} - F_{cont}),$$

$$\alpha = WN_0 E[a_{m_s,0}] + LF_{DL} P_{BS} \left(w + 6 E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] \right).$$

Ainsi nous appellerons capacité (sans *soft handover*) d'une session s demandant l'en-trée dans le réseau radio dans le sens descendant la valeur : $\alpha \cdot \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_s^{DL}} + w \nu_s^{DL} \gamma_s^{DL}}$. Puis si la somme des capacités des sessions actives et de la session s faisant la requête d'admission est inférieure à $DL_{threshold}$ alors la session s pourra être admise.

2.3.4 Equations avec *soft handover*

Lorsque les mobiles sont dans une zone de *soft handover*, ils sollicitent le service de plusieurs antennes dans le sens descendant, comme décrit à la section 1.3.3. Nous supposons que deux antennes desservent un mobile en *soft handover*. Ainsi, dans le sens descendant, de la même manière que dans le sens montant, la fiabilité du lien radio est améliorée en situation de *soft handover* grâce à la recombinaison des signaux issus des deux antennes dans le récepteur du mobile. En conséquence, le rapport $\frac{E_b}{N_t}$ nécessaire à la réception au mobile du signal émis par chaque antenne peut être plus faible que si le mobile n'était pas en *soft handover*. Holma et Toskala^[4] estiment que l'on a un gain de 2.7dB à 4dB sur la valeur de $\frac{E_b}{N_t}$ requise pour un usager à pied (3km/h), dépendant du

facteur d'atténuation entre lui et ses stations de base.

Notons SHG_{DL} le facteur de réduction de la valeur de $\frac{E_b}{N_t}$ requise pour un mobile en *soft handover*. Notons $p_{\text{SH}}^{\text{DL}}$ la proportion de sessions actives sur des mobiles en *soft handover* dans une cellule.

On note par la suite $P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s)$ l'expression de P_s^{DL} comme dépendant de la valeur donnée à γ_s^{DL} et de la position du mobile m_s :

$$P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) = \gamma_s^{\text{DL}} \frac{P_{\text{T}}^{\text{DL}} \left(w + \sum_{i \in \mathcal{B}} \frac{a_{m_s,0}}{a_{m_s,i}} \right) + N_0 W a_{m_s,0}}{\frac{W}{R_s^{\text{DL}}} + w \nu_s^{\text{DL}} \gamma_s^{\text{DL}}}. \quad (2.16)$$

Prenant en compte le *soft handover*, l'équation 2.11 devient:

$$P_{\text{T}}^{\text{DL}} = \sum_{s \in S^{\text{NSH}}} \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) + \sum_{s \in S^{\text{SH}}} \nu_s^{\text{DL}} P_s^{\text{DL}}(\text{SHG}_{\text{DL}} \gamma_s^{\text{DL}}, m_s) + P_{\text{cont}}$$

ce qui peut également s'écrire sous la forme:

$$P_{\text{T}}^{\text{DL}} = \sum_{s \in S} \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) + \sum_{s \in S^{\text{SH}}} \left(\nu_s^{\text{DL}} P_s^{\text{DL}}(\text{SHG}_{\text{DL}} \gamma_s^{\text{DL}}, m_s) - \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) \right) + P_{\text{cont}}. \quad (2.17)$$

On fait maintenant l'approximation suivante: on suppose que chaque mobile en *soft handover* est placé exactement au milieu entre les deux stations de base qui le desservent (voir l'annexe II). On note m^* un mobile situé à cet endroit. Lorsque l'on considère qu'il y a des mobiles en *soft handover*, nous considérons donc forcément les usagers répartis uniformément sur une surface circulaire de rayon $R > \frac{d_0}{2}$ où d_0 est la distance entre deux stations de base. En effet si $R < \frac{d_0}{2}$ il nous faut considérer qu'aucun mobile n'est pris en *soft handover*. L'équation (2.17) devient alors:

$$P_{\text{T}}^{\text{DL}} = \sum_{s \in S} \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) + \sum_{s \in S^{\text{SH}}} \left(\nu_s^{\text{DL}} P_s^{\text{DL}}(\text{SHG}_{\text{DL}} \gamma_s^{\text{DL}}, m^*) - \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m^*) \right) + P_{\text{cont}}.$$

Puis on suppose qu'il y a une fraction $p_{\text{SH}}^{\text{DL}}$ de toutes les sessions actives dans la cellule qui sont en *soft handover* (les usagers actifs étant répartis sur une surface circulaire

de rayon $R > \frac{d_0}{2}$). Donc après cette nouvelle approximation l'équation (2.17) devient finalement:

$$P_T^{\text{DL}} = \sum_{s \in S} \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_s) + p_{\text{SH}}^{\text{DL}} \sum_{s \in S} \left(\nu_s^{\text{DL}} P_s^{\text{DL}}(\text{SHG}_{\text{DL}} \gamma_s^{\text{DL}}, m_*) - \nu_s^{\text{DL}} P_s^{\text{DL}}(\gamma_s^{\text{DL}}, m_*) \right) + P_{\text{cont}}. \quad (2.18)$$

D'après les équations (2.16) et (2.18) on extrait l'expression finale de P_T^{DL} :

$$P_T^{\text{DL}} = \frac{P_{\text{cont}} N_0 W \sum_{s \in S} \gamma_s^{\text{DL}} \nu_s^{\text{DL}} \left(\frac{a_{m*,0} \text{SHG}_{\text{DL}} p_{\text{SH}}^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \frac{a_{m_s,0} - a_{m*,0} p_{\text{SH}}^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} \right)}{1 - \sum_{s \in S} \gamma_s^{\text{DL}} \nu_s^{\text{DL}} \left(\frac{(w + \sum_{i \in \mathcal{B}} \frac{a_{m*,0}}{a_{m*,i}}) \text{SHG}_{\text{DL}} p_{\text{SH}}^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \frac{(w + 6E[\frac{a_{m_s,0}}{a_{m_s,i}}]) - (w + \sum_{i \in \mathcal{B}} \frac{a_{m*,0}}{a_{m*,i}}) p_{\text{SH}}^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} \right)}. \quad (2.19)$$

On soumet cette dernière équation à la contrainte $P_T^{\text{DL}} < L F_{\text{DL}} \cdot P_{\text{BS}}$, et considérant que pour toute session s l'atténuation $a_{m_s,0}$ vaut $E[a_{m_s,0}]$ et le rapport $\frac{a_{m_s,0}}{a_{m_s,i}}$ vaut $E[\frac{a_{m_s,0}}{a_{m_s,i}}]$, on obtient alors la condition suivante:

$$\beta \sum_{s \in S} \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \delta \sum_{s \in S} \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{R_s^{\text{DL}}} + w \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} < D L_{\text{threshold}} \quad (2.20)$$

avec

$$\begin{aligned} D L_{\text{threshold}} &= P_{\text{BS}}(L F_{\text{DL}} - F_{\text{cont}}), \\ \beta &= p_{\text{SH}}^{\text{DL}} \text{SHG}_{\text{DL}} \left(N_0 W a_{m*,0} + L F_{\text{DL}} P_{\text{BS}} \left(w + \sum_{i \in \mathcal{B}} \frac{a_{m*,0}}{a_{m*,i}} \right) \right), \\ \delta &= N_0 W E[a_{m_s,0}] + L F_{\text{DL}} P_{\text{BS}} \left(w + 6E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] \right) \\ &\quad - p_{\text{SH}}^{\text{DL}} \left(N_0 W a_{m*,0} + L F_{\text{DL}} P_{\text{BS}} \left(w + \sum_{i \in \mathcal{B}} \frac{a_{m*,0}}{a_{m*,i}} \right) \right). \end{aligned}$$

Le calcul des coefficients $a_{m*,0}$ et $a_{m*,i}$ pour $i \in \mathcal{B}$ est détaillé dans l'annexe II. Remarquons de plus, que si l'on a $p_{\text{SH}}^{\text{DL}} = 0$, c'est-à-dire que l'on ne considère aucun mobile ayant un gain de macrodiversité, alors nous retrouvons l'équation (2.15) établie sans *soft handover*.

L'équation (2.20) devient donc l'équation d'admission avec le *soft handover* pris en

compte. On appellera alors la capacité d'une session s la valeur $\beta \frac{W}{R_s^{\text{DL}}} + \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \delta \frac{W}{R_s^{\text{UL}}} + \frac{\gamma_s^{\text{UL}} \nu_s^{\text{UL}}}{w \text{SHG}_{\text{UL}} \gamma_s^{\text{UL}} \nu_s^{\text{UL}}}$.

2.4 Validation des équations de capacité

2.4.1 Valeurs numériques des paramètres utilisés

TAB. 2.2 – Paramètres des équations de capacité radio

W	1.25×10^6 Hz	W	1.25×10^6 Hz
λ	0.55 (section 2.2.2)	N_0	1.2589×10^{-20} W/Hz
SHG_{UL}	0.54 (section 2.2.4)	P_{BS}	20 W (section 2.3.2)
LF_{UL}	0.6 (section 2.2.3)	F_{CONT}	0.2 (section 2.3.2)
ν_s^{UL}	voix: 0.67 ^[4] autres: 1	w	0.4 (section 2.3.2)
		LF_{DL}	0.6 (section 2.3.3)
		SHG_{DL}	0.54 (section 2.3.4)
		ν_s^{DL}	voix: 0.67 ^[4] autres: 1
		K	3600 (annexe II)

(a)

sens montant

(b)

sens descendant

2.4.2 Analyse des équations

Dans le sens montant, l'équation de capacité a été établie en considérant que les sessions actives dans la cellule sont initiées par des mobiles capables d'émettre assez de puissance pour être décodés à la station de base. Ainsi, la capacité de la cellule est limitée par la charge de trafic que l'on considère et est indépendante de la position des mobiles. De plus le modèle considère que les interférences inter-cellulaires sont une fraction fixe de la puissance totale reçue à la station de base. Or le rayon de couverture dans le sens montant est limité, étant donnée une charge de trafic, d'une part par la puissance maximum qu'un mobile peut émettre et par sa position, mais surtout par la position des mobiles des autres cellules qui créent des interférences inter-cellulaires, car suivant leur distance par rapport à la station de base de la cellule que l'on considère, les interférences inter-cellulaires seront plus ou moins importantes, et les performances de

la cellule considérée s'en trouveront changées. Voici donc une limitation du modèle dans le sens montant: les interférences inter-cellulaires ne prennent pas en compte la position des mobiles et les facteurs d'atténuation qui en découlent. De ce fait, ce modèle dans le sens montant n'est pas adapté pour faire un dimensionnement en termes de rayon de couverture. L'étude de nos équations dans le sens montant se borne donc à observer la capacité de la cellule en fonction de la charge de trafic qu'on lui impose, sachant que l'on est dans la situation où l'on a une fraction p_{SH}^{UL} de sessions en *soft handover*.

Dans le sens descendant, par contre, nous pouvons parler de rayon de couverture, c'est-à-dire, compte-tenu d'une certaine charge de trafic et de la position des mobiles, quelle est la surface que la station de base peut couvrir. En effet, la position des mobiles est prise en compte dans les équations. Pour mieux comprendre cette notion de couverture, examinons plus attentivement l'équation de capacité dans le sens descendant d'abord sans le *soft handover* pris en compte. Ainsi d'après l'équation (2.15) on a:

$$\alpha \sum_{s \in S} \frac{\nu_s^{DL} \gamma_s^{DL}}{\frac{W}{R_g^{DL}} + w \nu_s^{DL} \gamma_s^{DL}} < DL_{threshold}$$

avec $\alpha = WN_0 E[a_{m_s,0}] + LF_{DL} P_{BS} \left(w + 6. E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] \right).$

D'après l'annexe II, pour une valeur de R de 1000 m, $E[a_{m_s,0}] = 5.56 \times 10^7$; or le terme WN_0 est de l'ordre de 10^{-14} watts, donc on s'aperçoit, étant données les valeurs numériques des paramètres, que le terme de gauche dans l'expression de α est très négligeable pour des valeurs de R de dizaines ou de milliers de mètres. Ainsi on peut considérer que α dépend uniquement de la valeur prise par le rapport des atténuations $E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right]$. Or d'après les calculs faits dans l'annexe II, quelque soit la distribution uniforme considérée, ce rapport des atténuations dépend uniquement du rapport $\frac{R}{d_0}$, où R est le rayon de la surface sur laquelle on considère les usagers répartis et d_0 la distance entre deux stations de bases voisines. Cela est cohérent puisque nous tenons compte des interférences inter-cellulaires, donc si le rapport $\frac{R}{d_0}$ reste le même, on peut dilater le plan des cellules et le rapport des atténuations reste identique. Par contre, nous allons pouvoir étudier l'impact de la fluctuation du rapport $\frac{R}{d_0}$ sur la capacité de la cellule. Ainsi, faire

varier ce rapport s'interprète également en considérant que l'on a la distance entre les stations de base d_0 fixée et que l'on fait varier le rayon R de la surface sur laquelle on considère les usagers répartis.

Ainsi, si $R \leq \frac{d_0}{2}$ on est dans la configuration de la figure 2.1, où aucun mobile ne peut être considéré en *soft handover*.

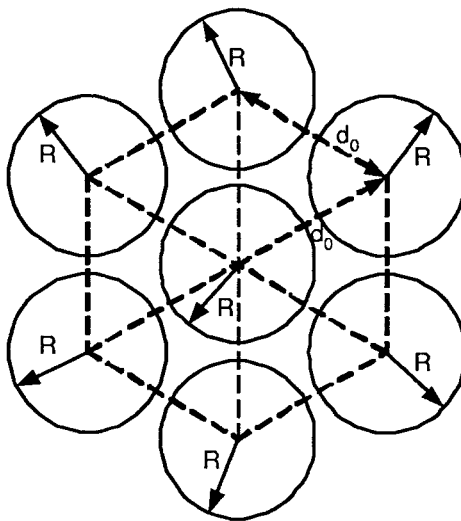


FIG. 2.1 – Configuration des cellules sans soft handover

Si $R > \frac{d_0}{2}$ on est alors dans la configuration de la figure 2.2, et il faut alors utiliser l'équation de capacité prenant en compte le *soft handover* (rappelons alors qu'une fraction p_{SH}^{DL} des sessions actives dans la surface de rayon R sont considérées comme étant en position de *soft handover*). En se référant à l'équation (2.20) et à l'annexe II on peut aisément vérifier que la capacité de la cellule ne dépend encore que du rapport $\frac{R}{d_0}$.

Maintenant il s'agit d'évaluer la proportion de mobiles en *soft handover* situés dans les zones de recouvrement "grisées" sur la figure 2.2. En effet, nous allons considérer que la proportion de mobiles en *soft handover* est égale à la proportion occupée en surface par les recouvrements par rapport à toute la surface de la cellule (ce qui est légitime puisque nous considérons que les usagers sont répartis uniformément dans la cellule).

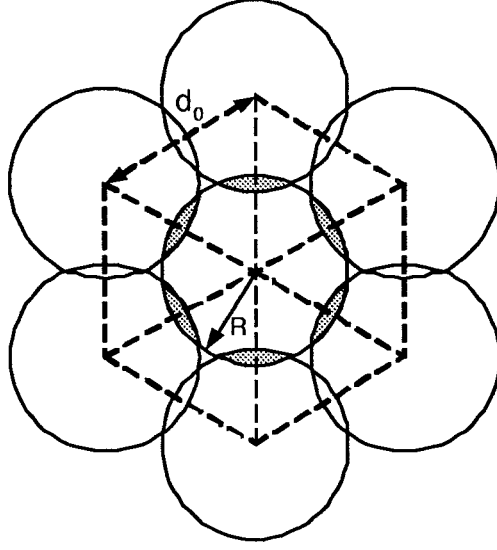


FIG. 2.2 – Configuration des cellules avec soft handover

Ainsi, géométriquement à partir de la figure 2.2, on a :

$$p_{\text{SH}}^{\text{DL}} \approx \frac{6}{\pi} \left(2 - \frac{d_0}{R} \right) \sqrt{1 - \frac{1}{4} \left(\frac{d_0}{R} \right)^2}. \quad (2.21)$$

La proportion de mobiles en *soft handover* dépend aussi uniquement du rapport $\frac{R}{d_0}$.

2.4.3 Influence du rayon de la cellule dans le sens descendant

Dans le sens descendant, les coefficients $a_{m_s,0}$ et $a_{m_s,i}$, $i \in \mathcal{B}$ sont calculés à partir des valeurs du rayon R de la surface où on suppose que les mobiles sont répartis, et de la distance d_0 séparant deux stations de bases voisines. D'autre part, comme nous l'avons vu dans la section précédente, la capacité d'une cellule dépend du type de trafic soumis et du rapport $\frac{R}{d_0}$. La figure 2.3 montre la capacité d'une cellule en nombre de sessions voix pouvant être admises en fonction du rapport $\frac{R}{d_0}$, soit en fonction de la surface sur laquelle sont réparties les sessions mobiles étant donné des stations de base fixées à une distance d_0 les unes des autres. Ainsi on utilise l'équation sans *soft handover* pour $\frac{R}{d_0} \leq 0.5$ et l'équation avec *soft handover* pour $\frac{R}{d_0} > 0.5$ afin de déterminer le nombre de sessions voix que la cellule est capable de contenir. On considère de la voix à 9600bps, un FER de

1%, et une distribution des usagers suivant la distance.

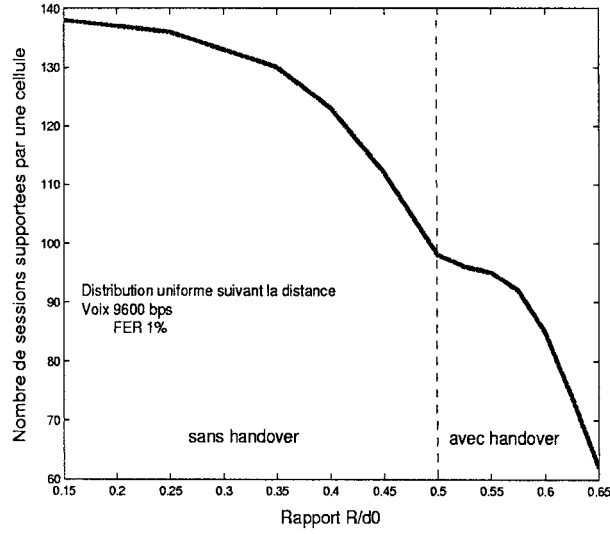


FIG. 2.3 – Influence du rayon

Il est normal d'observer une transition lorsque $R = \frac{d_0}{2}$. En effet, pour $\frac{R}{d_0} = 0.5$, nous considérons qu'aucune session ne bénéficie de la macrodiversité car aucune n'est en *soft handover*. Puis lorsque $R > \frac{d_0}{2}$, nous considérons qu'une fraction des sessions sur la surface de rayon R est en *soft handover* et donc pour cette fraction de sessions moins de puissance est nécessaire pour les satisfaire. Ainsi juste après $R = \frac{d_0}{2}$ on peut accepter à peu près le même nombre de sessions que pour $R = \frac{d_0}{2}$ en élargissant pourtant le rayon de la cellule. Avant et après cette phase de transition cette courbe montre bien que si l'on dispose de stations de base fixes équidistantes de d_0 , alors plus les usagers sont répartis uniformément sur une grande surface (soit dans nos équations, plus les mobiles sont situés sur un rayon moyen éloigné de la station de base), moins la station de base ne peut accepter de sessions.

2.4.4 Influence du type de trafic

Dans cette section, nous étudions la réaction des équations de capacité dans le sens montant et dans le sens descendant compte tenu du type de trafic généré dans la cellule. Dans cette étude on se place donc avec un rapport $\frac{R}{d_0}$ fixé dans le sens descendant, en se

mettant dans la configuration présentée sur la figure 2.4. Cette configuration des cellules dans le sens descendant est la configuration où l'on considère les mobiles répartis sur tout le plan, et où le *soft handover* est conséquent. En effet, dans cette configuration, $\frac{d_0}{R} = \sqrt{3}$ et $p_{SH}^{DL} = 25.6\%$. Pour rester cohérent on prend donc pour le sens montant $p_{SH}^{UL} = 25.6\%$. On rappelle qu'il s'agit du pourcentage de sessions reçues (sens descendant) ou initiées (sens montant) par un mobile en *soft handover*. Toute session est considérée comme étant bilatérale (un flot dans le sens montant et un flot dans le sens descendant), quitte à ce que l'un des flots soit de débit nul sur l'un des sens.

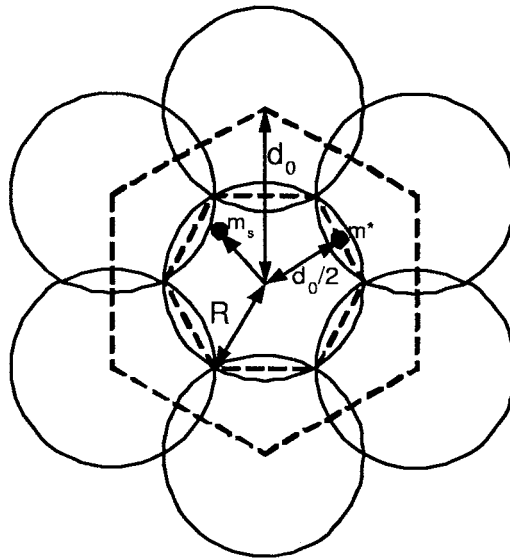


FIG. 2.4 – Géométrie des cellules

Nous allons considérer quatre chargements de trafic différents, afin d'évaluer l'impact du type de trafic sur la capacité de la cellule. Pour cela, le tableau 2.3 détaille les valeurs prises pour les RAB associés à chaque type d'application, dans le sens montant et dans le sens descendant. En effet, les trafics multimédias sont asymétriques.

TAB. 2.3 – RABs par types d'application

	voix	vidéo <i>streaming</i>	web	mail
UL:[débit ; fer]	[9.6kbps ; 1%]	[9.6kbps ; 1%]	[19.2kbps ; 5%]	[76.8kbps ; 5%]
DL:[débit ; fer]	[9.6kbps ; 1%]	[76.8kbps ; 1%]	[76.8kbps ; 5%]	[153.6kbps ; 5%]

Le tableau 2.4 quant à lui décrit les quatres types de trafic qui vont charger la cellule et être comparés. La figure 2.5 décrit, pour les trois chargements de trafic, le taux d'occupation à la station de base pour le sens montant (UL) et pour le sens descendant (DL). Les taux d'occupation à la station de base sont évalués de la façon suivante:

$$\text{taux d'occupation (\%)} = 100 \times \frac{\text{Capacité radio occupée}}{\text{Capacité radio totale disponible}}.$$

TAB. 2.4 – *Les différentes répartitions de trafic*

	voix	vidéo <i>streaming</i>	web	mail	chargement total (kbps)
trafic (a)	50 sessions	0 session	0 session	0 session	UL:480 DL:480
trafic (b)	25 sessions	0 session	2 sessions	0 session	UL:278.4 DL:393.6
trafic (c)	15 sessions	1 session	1 session	1 session	UL:249.6 DL:451.2
trafic (d)	2 sessions	2 sessions	2 sessions	2 sessions	UL:230.4 DL:633.6

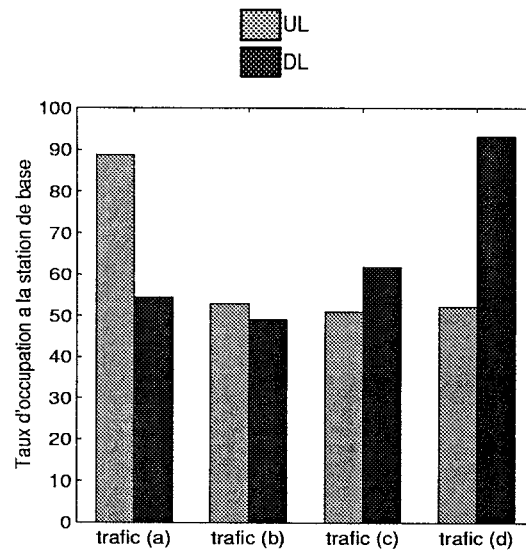


FIG. 2.5 – *Influence du type de trafic*

Ainsi, lorsque la cellule est chargée avec 50 sessions de voix (trafic (a)), le sens montant est plus rapidement saturé en termes de capacité que le sens descendant. Dans un contexte monoservice de type voix, le sens montant est le sens limitant la capacité de la cellule. Ceci s'explique du fait que le trafic de voix est symétrique et que les interférences intra-cellulaires sont plus importantes dans le sens montant que dans le sens descendant à cause de la non orthogonalité des codes. En revanche il apparaît au travers des char-

gements de trafic (b), (c), (d), que plus on introduit du trafic multimédia, plus le sens descendant devient le sens limitant la capacité de la cellule. En effet, le trafic des applications multimédias est fortement asymétrique et alors le sens descendant est le sens qui va congestionner le plus rapidement dans les réseaux de troisième génération, ce qui n'était pas le cas lorsque seulement la voix était utilisée.

2.5 Conclusion

Dans ce chapitre nous avons établi des formules de capacité radio pour le sens montant et pour le sens descendant ainsi que les bornes à ne pas dépasser pour que l'admission de la session entrante puisse se faire sans dégradation de qualité de service. L'analyse de ces équations met en évidence la notion de région de couverture dans le sens descendant et donc la possibilité d'envisager du dimensionnement à l'aide de ces formules d'admission. D'autre part, elles confirment le fait que le sens descendant est plus vite saturé que le sens montant dans le cas des futurs trafics multimédias fortement asymétriques. Ces validations sont nécessaires pour l'analyse des résultats de simulation qui sont faits dans le chapitre 4. De plus, nous avons pu établir les deux limitations suivantes dans le modèle développé.

- Dans le sens montant, les interférences inter-cellulaires sont une fraction fixe de la puissance totale reçue à la station de base. Donc, plus nous allons considérer de mobiles en *soft handover*, plus cette puissance totale sera petite grâce au gain en macrodiversité, et donc moins nous considérons d'interférences inter-cellulaires. Or, plus il y a de mobiles en *soft handover*, plus il y a d'interférences inter-cellulaires. Il s'agit donc de prendre une fraction de mobiles en *soft handover* pas trop importante pour que nos équations gardent un comportement proche de la réalité.
- Dans le sens descendant, nous avons mis en évidence la dépendance de la capacité de la cellule au rapport $\frac{R}{d_0}$. Ainsi, notre modèle suppose que les stations de base des différentes cellules sont réparties de manière équidistantes les unes des autres, à une distance d_0 . Cette distance étant fixée, chaque station de base voit son rayon de couverture diminuer lorsque la charge de trafic augmente. Cependant, nous suppo-

sons que toutes les cellules sont identiques en termes de trafic et de répartition des usagers, ce qui représente une limitation, puisque si le rayon de couverture d'une cellule diminue lorsqu'on lui augmente la charge de trafic, alors toutes les autres cellules sont supposées faire de même.

CHAPITRE 3

CONTRÔLE D'ADMISSION POUR LES RÉSEAUX 3G

3.1 Contexte d'admission

3.1.1 Modélisation du réseau

Dans le cadre de notre étude, nous considérons un réseau d'accès radio CDMA2000 avec un réseau cœur entièrement IP, tels que décrits dans les sections 1.2 et 1.3. Le réseau cœur est un réseau à intégration de services comme décrit à la section 1.4.3, s'appuyant sur le mode de commutation de paquets pour transmettre les données. Le protocole IP est le protocole choisi pour assurer la transmission bout en bout des paquets tant de voix que de données, comme décrit à la section 1.2.1.

Dans notre étude, nous supposons que tous les types d'application ont une demande de QoS, notamment une contrainte de délai à respecter, autant pour les applications temps réel que pour les applications non temps réel. Afin de permettre la garantie de qualité de service pour chaque flot, nous utilisons un mécanisme de réservation de ressources en chacun des nœuds du réseau appartenant au chemin de la session, par exemple le mécanisme RSVP décrit à la section 1.4.3. Cela suppose donc que pour tous les types d'application le mécanisme de réservation des ressources est utilisé, et notre étude émule donc le mode connecté pour toutes les sessions, tout en s'appuyant sur le protocole IP.

Nous rappelons que dans notre étude, la garantie de QoS concerne toute session initiée par un usager. Une session est la donnée d'une application et d'un profil d'utilisateur, comme décrit dans la section 1.4.2. Ainsi plusieurs applications initiées par un même usager sont considérées comme plusieurs sessions distinctes. Chaque flot, représentant une et une seule session, est traité de façon indépendante des autres flots. Une session peut donc être initiée par un terminal mobile et il s'agit alors d'une session dans le sens montant, soit elle est initiée par un terminal fixe et il s'agit alors d'une session dans le sens descendant.

Le service garanti aux sessions est assuré par l'ensemble des nœuds du réseau de

l'opérateur. Ainsi, conformément à la description du réseau CDMA2000 faite dans la section 1.2.1, une session bout en bout désigne dans notre étude les communications établies entre le terminal mobile et le nœud d'interconnexion avec le réseau externe.

La figure 3.1 représente la modélisation du réseau.

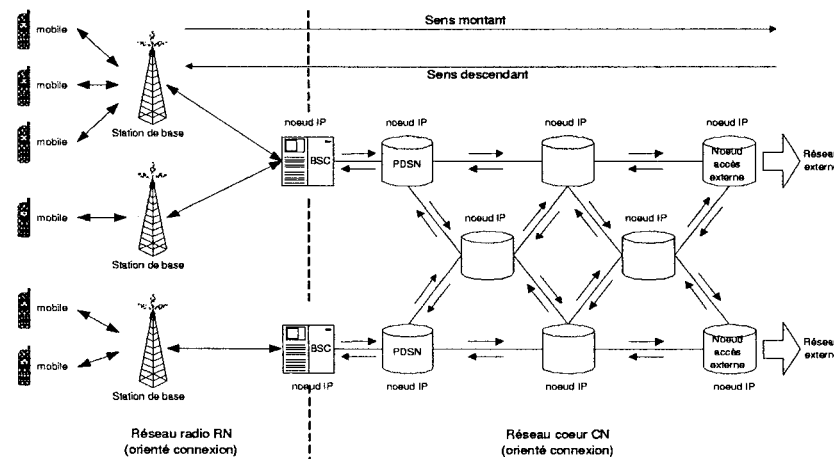


FIG. 3.1 – *Architecture du réseau modélisé*

3.1.2 Filtrage du trafic et discipline d'ordonnancement dans le réseau cœur

En chaque nœud du réseau cœur est implémentée la discipline de service WFQ décrite à la section 1.4.4, sous la condition RPPS décrite à la section 1.4.5.2. Afin d'utiliser les propriétés de cette discipline pour assurer une garantie pour chaque session du taux de service et du délai subi par un paquet, toute session s se voit allouer une seule file d'attente; puis en entrée du réseau cœur (sens montant et sens descendant) chaque file d'attente est précédée par un *token bucket* de paramètres $(\sigma_s, \rho_s^{\text{CN}})$ qui régule l'émission des paquets de la source initiant la session s . La profondeur et le taux de remplissage du *token bucket* pour une session acceptée dans le réseau sont fixés par la procédure de contrôle d'admission. Ainsi, conformément aux résultats de Pareck et Gallager présentés à la section 1.4.5.2, en dimensionnant adéquatement les tailles des files d'attentes selon l'expression 1.2, et en assurant à chaque nœud par le contrôle d'admission la condition 1.1 de stabilité du système, alors le réseau cœur est stable et en chaque nœud aucun phénomène de débordement des files d'attentes ne se produit. De plus, le débit moyen

de la session s sur chaque lien traversé est égal au taux de remplissage ρ_s^{CN} du *bucket* associé.

Remarquons que, dans le sens descendant, si la source de trafic initiant la session s en entrée du réseau cœur émet du trafic non conforme aux paramètres $(\sigma_s, \rho_s^{\text{CN}})$ de son *token bucket*, alors il peut y avoir des pertes de paquets dans les mémoires tampon des *buckets*. Une source conformante est une source qui émet ses paquets à un débit moyen ne dépassant pas ρ_s^{CN} et qui n'émet aucun burst de trafic de taille supérieure à σ_s . Ainsi une source non conformante dans le sens descendant engendre des pertes de paquets dans la mémoire tampon du *bucket* en entrée du réseau cœur, et de ce fait engendre une dégradation de la QoS offerte.

La figure 3.2 représente la modélisation des files d'attente du réseau de nœuds. Notons que les nœuds de périphérie BSC et nœuds d'accès externes sont plus complexes que les nœuds de cœur puisque le contrôle d'admission dans les nœuds de périphérie intègre le dimensionnement des *token buckets*.

3.1.3 Ressources sur l'interface radio

Sur le réseau d'accès radio, le contrôle d'admission se fait au niveau du BSC dans le sens montant et dans le sens descendant. Le contrôle d'admission adopté est celui présenté dans le chapitre 2, se basant sur la quantité d'interférences. Le BSC alloue les ressources radio pour chaque session s admise en lui allouant un RAB (*Radio Access Bearer*), c'est-à-dire en lui allouant un canal radio CDMA2000, conformément à la description faite dans la section 1.3.4. On rappelle qu'un RAB détermine le débit noté ρ_s^{RN} et le FER noté FER_s du canal alloué à la session s .

Remarquons que, dans le sens montant, si une source en entrée du réseau radio émet à un débit (que ce soit un débit moyen ou le débit d'un burst de trafic) supérieur au débit du RAB alloué, alors on assiste à des pertes de paquets en entrée du réseau et cela affecte donc la QoS offerte.

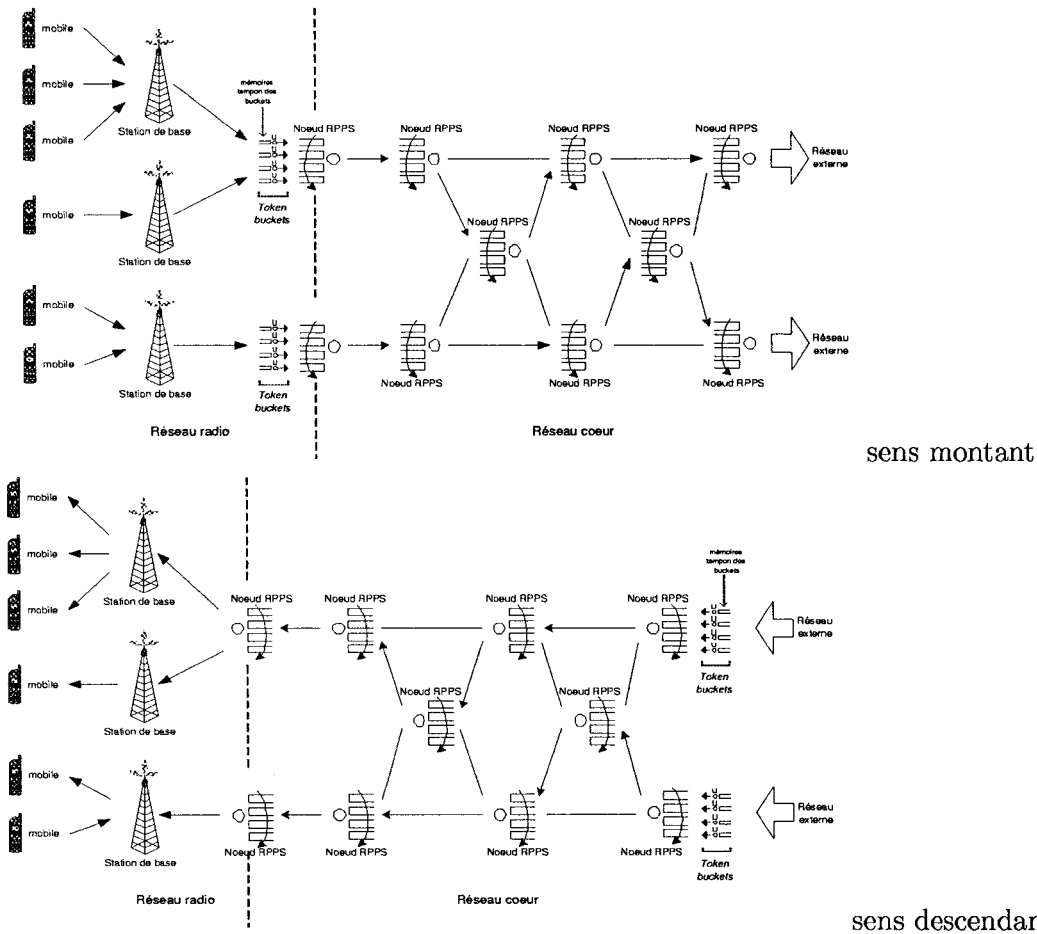


FIG. 3.2 – Modélisation des files d'attente

3.1.4 Composantes du délai bout en bout

3.1.4.1 Délai dans le réseau radio

Sur la partie radio (sens montant ou sens descendant), le délai expérimenté par les paquets d'une connexion est constitué du délai de transmission et du délai de propagation des paquets sur l'interface radio, ainsi que du délai de traitement (au niveau du terminal mobile et de la station de base). Les protocoles de transmission de données de l'interface radio CDMA2000 transmettent les données sur des trames de longueur fixe 20ms. Ainsi dans le sens descendant, au niveau de la station de base, un paquet IP est scindé pour être transmis par trames sur le lien radio. Dans le sens montant, les trames reçues à

la station de base sont réorganisées pour former un paquet IP à transmettre sur le lien câblé. On s'intéresse donc au délai subi par une trame sur le lien radio, puis on en déduira le délai subi par un paquet.

Au cours de sa transmission sur un canal radio-mobile, le signal subit différents types de détériorations pouvant se traduire par une dégradation de la qualité de la communication pour les usagers. Pour quantifier ces différentes formes de détérioration du signal, on utilise le taux d'erreur par trames (FER pour *Frame Error Rate*). On note ainsi que les transmissions en environnement radio-mobile doivent faire face de 10 000 à plus d'un milliard de fois plus d'erreurs que dans le cas des transmissions sur liaisons fixes. Par conséquent le réseau d'accès radio est conçu pour pouvoir fonctionner avec des taux d'erreurs par trames élevés, c'est-à-dire qu'il doit pouvoir corriger les erreurs en nombre important dans les cas des applications non temp-réel, car ces applications ne tolèrent pas les erreurs. Le contrôle d'erreur sur l'interface radio est notamment réalisé en utilisant la méthode de retransmission de l'ARQ (voir l'annexe III) qui corrige les erreurs en induisant un délai de retransmission pour chaque trame. Il n'est ainsi pas nécessaire de retransmettre les trames éronnées pour les applications temps réel car ces dernières tolèrent des erreurs mais doivent garder un faible délai sur le réseau.

Ainsi nous considérons le délai moyen subi par une trame sur le réseau d'accès radio. Sur l'interface radio, nous considérons qu'une trame qui ne subit pas de retransmission a un délai moyen incompressible $D_{tnr}=100\text{ms}$ dans le cas d'une trame non temps réel et de 40ms dans le cas d'une trame temps réel. On rappelle que ce délai prend en compte le délai de propagation, le délai de transmission de 20ms d'une trame et le délai de traitement. La différence entre le délai d'une trame temps réel et d'une trame non temps réel vient donc du délai de traitement qui diffère.

Afin de calculer le délai supplémentaire induit par les retransmissions pour les applications non temps réel, nous nous basons sur le protocole de répétition sélective décrit dans l'annexe III. On note D_{to} le temps de temporisation entre deux NACK successifs. Après l'envoi d'un premier NACK, l'émetteur doit attendre au moins un temps D_{to} pour renvoyer un nouveau NACK ($D_{to} > 2D_{tnr}$). Enfin on note D_t le délai moyen subi par une trame avec retransmissions, D_t^k le délai subi par une trame ayant subi k retransmissions

et P^k la probabilité pour une trame d'être retransmise k fois. On a ainsi, d'après la figure 3.3:

$$P^0 = 1 - \text{FER}$$

$$D_t^0 = D_{tnr}$$

$$P^1 = \text{FER}(1 - \text{FER})$$

$$D_t^1 = 3D_{tnr}$$

$$P^2 = \text{FER}^2(1 - \text{FER})$$

$$D_t^2 = 3D_{tnr} + D_{to}$$

$$P^3 = \text{FER}^3(1 - \text{FER})$$

$$D_t^3 = 3D_{tnr} + 2D_{to}$$

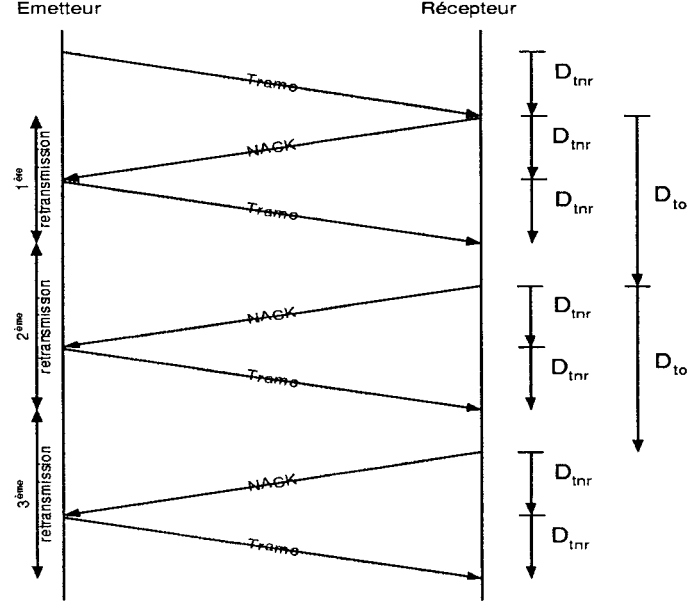


FIG. 3.3 – Délai de transmission radio

On considère qu'une trame n'a pas besoin de plus de trois retransmissions, et ainsi on obtient la valeur moyenne du délai d'une trame ayant subi des retransmissions:

$$\begin{aligned} D_t &= P^0 D_t^0 + P^1 D_t^1 + P^2 D_t^2 + P^3 D_t^3 \\ &= (1 - \text{FER}) D_{tnr} + \text{FER}(1 - \text{FER})(3D_{tnr}) + \text{FER}^2(1 - \text{FER})(3D_{tnr} + D_{to}) \\ &\quad + \text{FER}^3(1 - \text{FER})(3D_{tnr} + 2D_{to}). \end{aligned} \quad (3.1)$$

Ainsi le délai moyen $\overline{D}_{trame}$ subi par une trame sur le réseau d'accès radio est:

– $\overline{D}_{trame} = 40ms$ dans le cas d'une application temps réel,

– $\overline{D}_{trame} = 102.3$ ms pour un FER de 1%,

= 110.68 ms pour un FER de 5% , dans le cas d'une application non temps

réel.

Maintenant nous calculons le délai moyen subi par un paquet sur le réseau d'accès radio. Notons $L_s^{\max}$ la taille maximum d'un paquet d'une session s, ρ_s^{RN} le débit du RAB

alloué à la session s sur le lien radio (sens montant ou descendant) et T_{trame} la durée d'une trame ($T_{trame} = 20ms$). Alors le délai D_s^{RN} subi par un paquet de cette session sur le lien radio est:

$$D_s^{RN} = \overline{D}_{trame} \times \frac{L_s^{\max}}{T_{trame} \cdot \rho_s^{RN}}. \quad (3.2)$$

3.1.4.2 Délai dans le réseau cœur

Le délai expérimenté par un paquet sur les liens câblés du réseau cœur à commutation de paquets (sens montant ou sens descendant) est constitué de deux composantes distinctes: une partie fixe due au délai de propagation, au délai de transmission sur les liens et au délai de traitement au niveau des routeurs, et une partie variable due au délai d'attente dans les files d'attentes WFQ du réseau. Cette variabilité de la valeur du délai subi par un paquet est appelée gigue sur le délai (*delay jitter*) et le délai d'attente représente donc la composante du délai à contrôler afin de fournir des garanties de QoS aux applications, notamment aux applications temps réel. La figure 3.4 montre la composition du délai subi par un paquet sur le réseau cœur pour une session type.

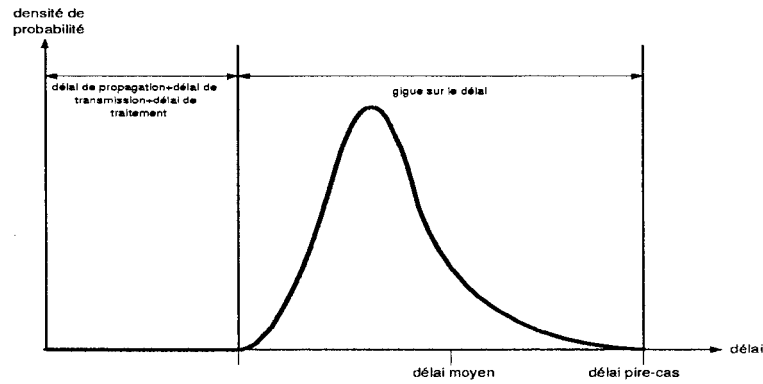


FIG. 3.4 – Composantes du délai d'un paquet transmis sur le réseau cœur

Notons D_s^{CN} le délai subi par un paquet de la session s sur le réseau cœur. Soit h_s le nombre de nœud sur le chemin supportant la session s .

Soit $D_{s,m}^p$ et $D_{s,m}^{tr}$ respectivement le temps de propagation et le temps de transmission sur le lien entre les nœuds consécutifs m et $m + 1$. Alors le délai de propagation total D_s^p le long du chemin supportant la session s dépend de la longueur des liens et vaut:

$$D_s^p = \sum_{m=1}^{h_s} D_{s,m}^p.$$

Le délai de transmission total D_s^{tr} le long du chemin supportant la session s dépend de la taille des paquets de la session et de la bande passante allouée à la session s et vaut:

$$D_s^{tr} = \sum_{m=1}^{h_s} D_{s,m}^{tr}.$$

Soit $\overline{D}_s^t$ le temps de traitement moyen subi par les paquets appartenant à la session s en chacun des nœuds m du chemin emprunté par la session. Le temps de traitement total D_s^t est donc:

$$D_s^t = h_s \overline{D}_s^t.$$

Soit D_s^q le temps d'attente d'un paquet de la session s dans les files d'attente sur le chemin emprunté. On a alors la décomposition du délai suivante:

$$D_s^{CN} = D_s^q + D_s^t + D_s^p + D_s^{tr}. \quad (3.3)$$

3.1.4.3 Délai bout en bout

Notons D_s le délai bout en bout (sens montant ou sens descendant) subi par un paquet d'une session s sur le réseau entre le terminal mobile et le nœud d'accès externe. On a donc:

$$\begin{aligned} D_s &= D_s^{CN} + D_s^{RN} \\ &= D_s^q + D_s^t + D_s^p + D_s^{tr} + \overline{D}_{trame} \times \frac{L_s^{\max}}{T_{trame} \cdot \rho_s^{RN}}. \end{aligned} \quad (3.4)$$

Pour une session s donnée seule la composante D_s^q du délai d'attente dans le réseau cœur est variable. Toutes les autres composantes sont fixes. Or, dans notre étude, nous utilisons les propriétés de la discipline de service WFQ sous la condition RPPS présentées dans la section 1.4.5.2 afin de contrôler ce délai d'attente. Le contrôle sur la valeur du délai d'attente pour garantir le délai bout en bout est décrit dans la section 3.1.5

3.1.5 Garantie de délai et de bande passante

Dans le cadre des réseaux de troisième génération, avec la présence à la fois de sessions temps réel et non temps réel, la contrainte de délai est une exigence de QoS essentielle. L'algorithme d'admission se donne pour mission de limiter le délai d'attente dans les files des paquets appartenant à une même session afin de contrôler le délai bout en bout maximal subi par les paquets dans le réseau en entier. De plus, la procédure d'admission doit permettre de respecter aussi les contraintes exprimées par la source en termes de débit, en allouant des taux de service adéquats en chacun des nœuds du réseau cœur, et des RABS adéquats dans le réseau radio pour que le débit sur la partie radio soit garanti.

Afin de garantir un certain débit pour une session acceptée dans le réseau (sens montant ou sens descendant), l'algorithme d'admission utilise pour la partie réseau cœur la propriété de WFQ sous la condition RPPS qui garantit qu'en chaque nœud m sur le chemin de la session s , le taux de service de la session s est au moins égal au taux de remplissage du *token bucket*. Précisons que cela est vrai car le contrôle d'admission a au préalable garanti la condition 1.1 en acceptant la session dans le réseau. Ainsi, si r_s est la demande en bande passante d'une session s faisant sa requête au réseau (sens montant ou sens descendant), alors il suffit de fixer le taux de remplissage du *bucket* associé à cette session à $\rho_s = r_s$. En effet, on est alors assuré sur la partie réseau cœur que la session s est servie en chaque nœud à un débit au moins égal à r_s . Sur la partie radio, il suffit d'allouer à la session s le RAB dont le débit est supérieur ou égal à la requête r_s .

Afin de garantir le délai bout en bout subi par un paquet d'une session acceptée dans le réseau (sens montant ou sens descendant), l'algorithme d'admission utilise les propriétés de la discipline d'ordonnancement WFQ sous la condition RPPS qui offre une borne $D_s^{q, \max}$ sur le délai d'attente maximal des paquets d'une session s donnée. Précisons que cela est vrai car le contrôle d'admission a au préalable garanti la condition 1.1 en acceptant la session dans le réseau. Soit $D_s^{\max}$ la contrainte de délai bout en bout (sens montant ou sens descendant) exprimée par la session s lors de sa requête au réseau et δD_s la variation sur le délai tolérée. Alors, pour satisfaire cette contrainte il suffit que la borne $D_s^{q, \max}$ soit égale à:

$$D_s^{q, \max} = D_s^{\max} + \delta D_s - (D_s^t + D_s^p + D_s^{tr} + \bar{D}_{trame} \times \frac{L_s^{\max}}{T_{trame} \cdot \rho_s^{\text{RN}}}). \quad (3.5)$$

Pour garantir cette borne sur le délai d'attente, il s'agit donc d'après l'équation (1.3) établie dans la section 1.4.5.2 de dimensionner correctement la profondeur du *token bucket* associé à la session s . En effet, la borne établie par Pareck et Gallager dépend des paramètres du *token bucket* associé à la session s . Or, le taux de remplissage ρ_s est fixé en fonction de la demande en bande passante afin de garantir le débit pour la session s , donc le paramètre à dimensionner pour garantir le délai est la profondeur du *bucket*. On obtient ainsi d'après l'équation (1.3) la profondeur à fixer pour assurer que le délai dans les files d'attente ne dépasse pas $D_s^{q, \max}$ et donc que le délai bout en bout ne dépasse pas $D_s^{\max} + \delta D_s$:

$$\sigma_s = \rho_s \left(D_s^{q, \max} - \sum_{m=1}^{h_s} \frac{L_{\max}}{r^m} \right) - 2(h_s - 1)L_s^{\max} \quad (3.6)$$

avec la valeur de $D_s^{q, \max}$ calculée à partir de l'équation (3.5).

3.1.6 Spécification de QdS

3.1.6.1 Descripteurs de QdS des sessions demandant l'accès au réseau

Dans le cadre du service garanti défini à la section 1.6.1, les paramètres décrivant les sources de trafic sont nécessaires à la caractérisation à priori de ces sources. Ainsi une session se présentant à l'entrée du réseau (sens montant ou sens descendant) a une demande en termes de débit et la source initiant cette session s'engage à transmettre conformément au débit demandé. Dans notre étude, une session s demande une garantie de débit dans l'intervalle $[r_s^{\min}, r_s^{\max}]$, $r_s^{\max}$ désignant le débit moyen maximal requis par la source et $r_s^{\min}$ désignant le débit moyen minimal tolérable par la source. Cet intervalle permettra plus tard de modifier le débit alloué à une session dans le réseau et rendre le contrôle d'admission adaptatif. On suppose donc par la suite que la source (sens montant ou sens descendant) émet du trafic conforme à ses déclarations, sous peine de se voir attribuer une QdS en termes de débit qui ne la satisferait pas. La figure 3.5 représente le débit réel envoyé par la source de la session s ainsi que son intervalle de demande de

QdS $[r_s^{\min}, r_s^{\max}]$, et un un exemple d'évolution que peut subir le débit alloué à la session s . La valeur de $r_s^{\min}$ représente le débit minimum que la session peut tolérer, c'est-à-dire que la QdS en termes de débit reste satisfaisante si ce débit doit être alloué à la session.

débit de transmission

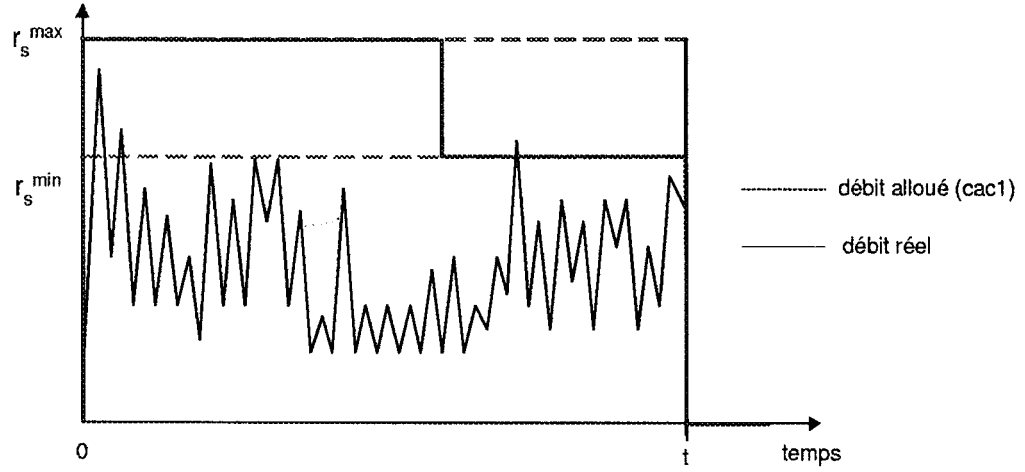


FIG. 3.5 – Evolution du débit alloué à la session s lors de sa période d'activité $[0, t)$ sous la procédure CAC1

En plus de ces descripteurs de trafic nécessaires à la caractérisation du trafic attendu, une session faisant une requête au réseau a des exigences de QdS telles que le délai par paquet maximal toléré $D_s^{\max}$, la gigue tolérable sur le délai δD_s , la priorité de l'utilisateur P_s selon la hiérarchie établie à la section 1.4.2, et le taux d'erreur demandé sur la partie radio F_s . Ces exigences de QdS sont regroupées dans le tableau 3.1.

TAB. 3.1 – Exigences de QdS d'une session faisant une requête au réseau

$r_s^{\max}$	débit maximal requis
$r_s^{\min}$	débit minimum tolérable
$D_s^{\max}$	délai par paquet maximal toléré
δD_s	gigue tolérable sur le délai
P_s	priorité de l'utilisateur
F_s	FER requis

3.1.6.2 Descripteurs de trafic des sessions en cours

L'algorithme d'admission garantit que toute session acceptée dans le réseau a ses exigences en termes de QoS satisfaites pendant toute sa durée de vie. Cela suppose donc le maintien d'états par session en chacun des nœuds du réseau cœur et sur l'interface radio. Ces états contiennent les paramètres nécessaires à la transmission des paquets sur le réseau ainsi qu'au contrôle de flot. Ces états sont à rapprocher des états du mécanisme de réservation RSVP décrit dans la section 1.4.3: les exigences de QoS décrites dans le tableau 3.1 et les caractéristiques du chemin emprunté par une session nécessaires à l'algorithme d'admission (nombre de nœuds traversés h_s et la somme des capacités r^m des liens empruntés par la session) sont contenues dans la structure *TSpec*. Lorsqu'une session est admise dans le réseau, les ressources sont réservées sur tout le chemin de la session et les descripteurs de trafic des sessions admises dans le réseau sont contenus dans la structure *RSpec*, les paramètres contenus dans cette structure pouvant être actualisés pendant la durée de vie d'une session, ce qui permettra par la suite d'établir un contrôle de flot.

Ainsi, toute session admise dans le réseau (sens montant ou sens descendant) est entièrement décrite par les paramètres $(\sigma_s, \rho_s^{\text{CN}})$ alloués à son *token bucket* pour la partie de cœur, et par le RAB RAB_s alloué sur la partie radio. Rappelons que ce RAB est composé du débit du canal radio alloué à la session s , noté ρ_s^{RN} , et du taux d'erreur par trame FER_s du canal alloué. Le tableau 3.2 regroupe ces descripteurs de trafic des sessions admises dans le réseau.

TAB. 3.2 – *Descripteurs de trafic*

ρ_s^{CN}	taux de remplissage du bucket
σ_s	profondeur du bucket
FER_s	taux d'erreur subi sur la partie radio
ρ_s^{RN}	débit du canal radio alloué

Ainsi, que ce soit dans le sens montant ou dans le sens descendant, ρ_s^{CN} représente la bande passante allouée et réservée sur chaque lien emprunté par la session s sur la partie réseau cœur, et ρ_s^{RN} représente la bande passante allouée sur la partie radio. FER_s est le

taux d'erreur par trames subi par la session s sur la partie radio et détermine le délai subi sur l'interface radio. Quant au paramètre σ_s , il représente la profondeur du *bucket* associé à la session s et il détermine également la taille (en termes de mémoire tampon nécessaire) de la file d'attente de la session en chaque nœud du chemin. En effet, d'après l'équation (1.2), la quantité maximale de trafic qui peut s'accumuler dans la file d'attente de la session en un nœud quelconque du chemin emprunté est $\sigma_s + L_s^{\max}$, avec $L_s^{\max}$ qui est la taille maximale d'un paquet de la session s . Ainsi, $\sigma_s + L_s^{\max}$ est la taille allouée en chaque nœud du chemin de la session s à la mémoire tampon de sa file d'attente.

3.2 Contrôle d'admission pour du service garanti

3.2.1 Notations

Dans cette étude, une session bilatérale qui met en jeu un flot montant et un flot descendant est considérée comme étant composée de deux sessions distinctes, l'une sur le sens montant, l'autre sur le sens descendant. Ainsi, la session sur le sens montant est considérée comme étant logiquement indépendante de la session sur le sens descendant: les chemins suivis et les allocations de ressources sont faits indépendamment. De ce fait, l'admission dans le réseau d'une session bilatérale ne se fait que si la session sur le lien montant et la session sur le lien descendant peuvent être admises. Par la suite nous considérons donc les conditions d'admission d'une session s individuelle sur l'un des sens (montant ou descendant). Les notations utilisées par la suite sont les suivantes:

s	indice de la session faisant la requête d'admission au réseau.
$L_s^{\max}$	longueur maximale d'un paquet de la session s .
h_s	nombre de nœuds sur le chemin devant supporter la session s .
m	indice d'un nœud sur le chemin devant supporter la session s .
$\mathcal{P}(s)$	ensemble des nœuds du chemin devant supporter la session s : $\mathcal{P}(s) = \{1, 2, \dots, h_s\}$.
$\mathcal{L}_s^m$	lien de sortie au nœud m appartenant au chemin de la session s .
r^m	capacité du lien $\mathcal{L}_s^m$ au nœud m .

$\mathcal{B}^m$	espace mémoire maximal disponible au nœud m pour le lien $\mathcal{L}_s^m$.
s'	indice d'une session active en cours de transmission dans le réseau.
$\mathcal{S}_m$	ensemble des sessions actives au nœud m sur le lien $\mathcal{L}_s^m$ dans le réseau cœur.
$\mathcal{S}$	ensemble des sessions actives sur l'interface radio dans la même cellule que s .
$\Delta_{s'}^m$	longueur de la file d'attente d'une session s' au nœud m .
$\rho_{s'}^{\text{CN}}$	débit d'une session s' sur le réseau cœur.
$\rho_{s'}^{\text{RN}}$	débit du RAB d'une session s' sur le réseau radio.
$\sigma_{s'}$	profondeur du <i>bucket</i> d'une session s' .
$\text{FER}_{s'}$	FER d'une session s' sur l'interface radio.
$\mathcal{C}_{\text{UL}}(s')$	capacité radio utilisée par une session s' si sens montant.
$\mathcal{C}_{\text{DL}}(s')$	capacité radio utilisée par une session s' si sens descendant.
ρ_s^{CN}	débit alloué à la session s sur le réseau cœur (taux de remplissage du <i>bucket</i>).
ρ_s^{RN}	débit du RAB alloué à la session s sur le réseau radio.
σ_s	profondeur du <i>bucket</i> allouée à la session s .
Δ_s^m	taille de la file d'attente allouée à la session s au nœud m .
FER_s	FER alloué à la session s sur l'interface radio.
$\mathcal{C}_{\text{UL}}(s)$	capacité radio nécessaire à la session s si sens montant.
$\mathcal{C}_{\text{DL}}(s)$	capacité radio nécessaire à la session s si sens descendant.

Rappelons que les exigences de QdS de la session s sont récapitulées dans le tableau 3.1 dont on tire les notations.

Par la suite, on note $\mathcal{R}(r_s)$ le débit du RAB du canal radio dont la valeur est immédiatement supérieure à la valeur de débit r_s . En effet, sur la partie radio, les débits disponibles sont fixés par la configuration radio choisie et prennent des valeurs discrètes détaillée à la section 1.3.4. Conformément aux configurations radio que nous avons choisies ces valeurs de débit sont les mêmes dans le sens montant et dans le sens descendant.

3.2.2 Vérification des ressources disponibles

Toute session s demande à être admise dans le réseau au débit de transmission maximal $r_s^{\max}$. Puis pendant la période d'activité de la session, d'éventuelles renégociations auront lieu afin de modifier le débit alloué dans l'intervalle $[r_s^{\min}, r_s^{\max}]$.

Sens montant. Soit une session s faisant une requête au réseau dans le sens montant. Le nœud $m = 1$ est le nœud BSC et le nœud $m = h_s$ est le nœud d'accès externe. Le terminal mobile envoie un message de signalisation supportant les exigences de QdS de la session vers sa station de base. Le nœud BSC enregistre et transmet ces spécifications de QdS sur le réseau cœur via un message de signalisation supportant la structure $TSpec_s$ du mécanisme RSVP, du nœud 1 au nœud destination h_s afin de déterminer les caractéristiques du chemin emprunté $\mathcal{P}(s)$. En chacun des nœuds du chemin $\mathcal{P}(s)$, la structure $TSpec_s$ met à jour les champs nécessaires à la détermination des paramètres h_s et $\sum_{m=1}^{h_s} \frac{1}{r^m}$, paramètres qui servent ensuite dans l'algorithme d'admission.

Ainsi avec la connaissance des paramètres descriptifs du chemin $\mathcal{P}(s)$ de la session s et de ses exigences de QdS, chaque nœud du réseau évalue si le montant des ressources nécessaires à la session est inférieur au montant des ressources disponibles. Le premier nœud BSC $m = 1$ évalue les ressources radio disponibles dans le sens montant et rend une décision positive ou négative sur l'admission dans le réseau radio. Ce même nœud BSC $m = 1$ et les nœuds $m=2, \dots, h_s - 1$ évaluent les ressources disponibles sur le liens câblés et rendent chacun une décision positive ou négative sur l'admission en chacun de ces nœuds du réseau cœur. Si la décision d'admission est positive sur le réseau radio et en chaque nœud du réseau cœur, alors la session s est admise dans le réseau et les ressources sont allouées et réservées de bout en bout.

Sens descendant. Soit une session s faisant une requête au réseau dans le sens descendant. Le nœud $m = 1$ est le nœud d'accès externe et le nœud $m = h_s$ est le nœud BSC. Le nœud d'entrée $m = 1$ transmet sur le réseau cœur un message de signalisation supportant la structure $TSpec_s$ du mécanisme RSVP, contenant les spécifications de QdS, du nœud 1 au nœud destination h_s afin de déterminer les caractéristiques du chemin emprunté $\mathcal{P}(s)$. En chacun des nœuds du chemin $\mathcal{P}(s)$, la structure $TSpec_s$ met à jour les

champs nécessaires à la détermination des paramètres h_s et $\sum_{m=1}^{h_s} \frac{1}{r^m}$, paramètres qui servent ensuite dans l'algorithme d'admission.

Ainsi avec la connaissance des paramètres descriptifs du chemin $\mathcal{P}(s)$ de la session s et de ses exigences de QoS, chaque nœud du réseau évalue si le montant des ressources nécessaires à la session est inférieur au montant des ressources disponibles. Les nœuds $m = 1, \dots, h_s - 1$ évaluent les ressources disponibles sur les liens câblés et rendent chacun une décision positive ou négative sur l'admission en chacun de ces nœuds du réseau cœur. Le dernier nœud BSC $m = h_s$, quant à lui, évalue les ressources radio disponibles dans le sens descendant et rend une décision positive ou négative sur l'admission dans le réseau radio. Si la décision d'admission est positive en chaque nœud du réseau cœur et sur l'interface radio, alors la session s est admise dans le réseau et les ressources sont allouées et réservées de bout-en-bout.

3.2.2.1 Test d'admission sur l'interface radio

Le test d'admission fait par le nœud BSC ($m = 1$ dans le sens montant et $m = h_s$ dans le sens descendant) pour la décision d'admission sur l'interface radio vient de l'équation de capacité radio (2.8) établie à la section 2.2.4 pour le sens montant et de l'équation (2.20) établie dans la section 2.3.4 pour le sens descendant:

Sens montant.

$$C_{UL}(s) + \sum_{s' \in \mathcal{S}} C_{UL}(s') < LF_{UL} \quad (3.7)$$

avec

$$C_{UL}(s) = \frac{1 + \lambda}{W} \frac{\gamma_s^{UL} \nu_s^{UL}}{\frac{1}{(1 - p_{SH}^{UL}(1 - SHG_{UL}))\mathcal{R}(r_s^{\max})} + \frac{\gamma_s^{UL} \nu_s^{UL}}{W}}.$$

Rappelons que γ_s^{UL} dépend de la requête de FER F_s et du débit $\mathcal{R}(r_s^{\max})$ comme détaillé dans la section 2.2.

$$C_{UL}(s') = \frac{1 + \lambda}{W} \frac{\gamma_{s'}^{UL} \nu_{s'}^{UL}}{\frac{1}{(1 - p_{SH}^{UL}(1 - SHG_{UL}))\rho_{s'}^{RN}} + \frac{\gamma_{s'}^{UL} \nu_{s'}^{UL}}{W}}$$

sachant que $\gamma_{s'}^{UL}$ dépend de $FER_{s'}$ et de $\rho_{s'}^{RN}$.

Sens descendant.

$$\mathcal{C}_{\text{DL}}(s) + \sum_{s' \in \mathcal{S}} \mathcal{C}_{\text{DL}}(s') < DL_{\text{threshold}} \quad (3.8)$$

avec

$$\mathcal{C}_{\text{DL}}(s) = \beta \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{\mathcal{R}(r_s^{\text{max}})} + w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \delta \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{\mathcal{R}(r_s^{\text{max}})} + w \gamma_s^{\text{DL}} \nu_s^{\text{DL}}}.$$

Rappelons que γ_s^{DL} dépend de la requête de FER F_s et du débit $\mathcal{R}(r_s^{\text{max}})$ comme détaillé dans la section 2.3.

$$\mathcal{C}_{\text{DL}}(s') = \beta \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{\rho_{s'}^{\text{RN}}} + w \text{SHG}_{\text{DL}} \gamma_s^{\text{DL}} \nu_s^{\text{DL}}} + \delta \frac{\gamma_s^{\text{DL}} \nu_s^{\text{DL}}}{\frac{W}{\rho_{s'}^{\text{RN}}} + w \gamma_s^{\text{DL}} \nu_s^{\text{DL}}}$$

sachant que $\gamma_{s'}^{\text{DL}}$ dépend de $\text{FER}_{s'}$ et de $\rho_{s'}^{\text{RN}}$.

3.2.2.2 Test d'admission sur les nœuds du réseau cœur

Le test d'admission sur le réseau cœur (sens montant ou sens descendant) est fait par tout nœud $m \in \{1, \dots, h_s - 1\}$:

$$r_s^{\text{max}} + \sum_{s' \in \mathcal{S}_m} \rho_{s'}^{\text{CN}} < r^m, \quad (3.9)$$

$$\sigma_s^{\text{max}} + L_s^{\text{max}} + \sum_{s' \in \mathcal{S}_m} \Delta_{s'}^m < \mathcal{B}^m \quad (3.10)$$

avec $\Delta_{s'}^m = \sigma_{s'} + L_{s'}^{\text{max}}$ et

$$\sigma_s^{\text{max}} = r_s^{\text{max}} \left(D_s^{q, \text{max}} - \sum_{m=1}^{h_s} \frac{L_{\text{max}}}{r^m} \right) - 2(h_s - 1)L_s^{\text{max}} + L_s^{\text{max}},$$

$D_s^{q, \text{max}}$ étant calculée à partir de l'équation (3.5).

3.2.3 Réserve et allocation des ressources

Toute session s admise dans le réseau, que ce soit dans le sens montant ou dans le sens descendant se voit attribuer les ressources suivantes lors de son admission:

$$\begin{aligned}\rho_s^{\text{CN}} &:= r_s^{\text{max}}, \\ \sigma_s &:= \sigma_s^{\text{max}}, \\ \Delta_s^m &:= \sigma_s^{\text{max}} + L_s^{\text{max}} \quad \forall m \in \{1, \dots, h_s - 1\}, \\ \rho_s^{\text{RN}} &:= \mathcal{R}(r_s^{\text{max}}), \\ \text{FER}_s &:= F_s.\end{aligned}$$

On rappelle que la source initiant la session s (sens montant ou sens descendant) émet un trafic conformant à ses exigences de QdS, car si cela n'était pas le cas, la QdS offerte à la session en termes de débit ne serait pas satisfaisante.

Par définition, on a $\rho_s^{\text{RN}} \geq \rho_s^{\text{CN}}$. Dans le sens descendant, cela assure qu'il n'y aura pas d'accumulation de paquets en entrée du réseau radio. Dans le sens montant, le trafic émis par la source est transmis dans le canal radio avant d'arriver au réseau cœur; si la source émet conformément à ses exigences de QdS en termes de débit, ce que nous considérons dans notre étude, alors il n'est pas grave que le débit alloué sur la partie radio soit supérieur au débit alloué sur le réseau cœur, le lien radio étant simplement sous-utilisé. Si la source n'est plus conformante, cas pouvant arriver dans la réalité, alors on peut assister à des pertes de paquets à l'entrée du réseau cœur dans le sens montant.

3.2.4 Renégociation de QdS

Nous nous plaçons dans le cas où l'un au moins des tests d'admission 3.7 (si sens montant), 3.8 (si sens descendant), 3.9 et 3.10 a échoué. Dans ce cas nous sommes dans un état de congestion, et il n'y a pas suffisamment de ressources dans le réseau pour accepter la session s qui se présente en entrée. La politique d'admission consistant à refuser la session s à ce moment là est nommée CAC0. Une autre politique adaptative, et nommée CAC1, consiste à libérer des ressources allouées aux sessions en cours de plus

faible priorité afin d'éventuellement pouvoir admettre la session entrante de plus haute priorité. Cela implique une renégociation de la QoS offerte aux sessions en cours s' en termes de débit, ce dernier demeurant cependant dans l'intervalle de requête $[r_{s'}^{\min}, r_{s'}^{\max}]$, tout en continuant de garantir le délai par paquet exigé par ces sessions.

Les pseudo-codes des algorithmes CAC0 et CAC1 sont présentés dans l'annexe IV.

3.2.4.1 Congestion sur l'interface radio

Si la condition 3.7 (si sens montant) ou la condition 3.8 (si sens descendant) n'est pas satisfaite, alors il y a congestion sur l'interface radio. Les sessions $s' \in \mathcal{S}$ sont alors classées dans l'ordre de priorité croissante. Puis, tant que le test d'admission sur la partie radio n'est pas satisfait, les sessions $s' \in \mathcal{S}$ telles que $P_{s'} < P_s$ font l'objet, une à une, par ordre croissant de priorité, d'une désallocation de leur montant de ressources allouées si cela est possible, c'est-à-dire si $\rho_{s'}^{\text{RN}} > \mathcal{R}(r_s^{\min})$. Pour une telle session $s' \in \mathcal{S}$ la désallocation effectuée est la suivante:

$$\begin{aligned}\rho_{s'}^{\text{RN}} &:= \mathcal{R}(r_{s'}^{\min}), \\ \rho_{s'}^{\text{CN}} &:= r_{s'}^{\min}, \\ \sigma_{s'} &:= \sigma_{s'}^{\min}, \\ \Delta_{s'}^m &:= \sigma_{s'}^{\min} + L_{s'}^{\max}\end{aligned}$$

avec

$$\sigma_{s'}^{\min} = r_{s'}^{\min} \left(D_{s'}^{q, \max} - \sum_{m=1}^{h_{s'}} \frac{L_{\max}}{r^m} \right) - 2(h_{s'} - 1)L_{s'}^{\max} + L_{s'}^{\max},$$

$D_{s'}^{q, \max}$ étant calculée à partir de l'équation (3.5).

Ainsi, sous condition que les ressources pouvant être libérées soient suffisantes, la session entrante s est acceptée dans le réseau et les ressources des sessions s' concernées sont libérées.

Suite à l'apparition d'un phénomène de congestion sur l'interface radio (sens montant ou sens descendant) lors de l'admission d'une session s , les montants de ressources allouées aux sessions s' en cours dans la même cellule que s font éventuellement l'objet d'une

réévaluation. Cette réévaluation du montant des ressources d'une session s' en cours est répercutée en chacun des nœuds du chemin $\mathcal{P}(s')$ sur le réseau cœur en mettant à jour les ressources allouées dans la structure $RSpec$ du mécanisme RSVP. Le contrôle de flot est donc effectué par la modification des $Rspec$ des sessions dégradées.

Remarquons que toute session s' voyant sa bande passante dégradée conserve par contre le même taux d'erreur par trames $FER_{s'}$ que celui demandé à l'entrée dans le réseau. Ainsi la puissance que doit émettre le mobile (si sens montant) ou la station de base (si sens descendant) initiant s' diminue et cela libère de la capacité radio pour la session demandant l'accès au réseau.

3.2.4.2 Congestion en un nœud du réseau cœur

Si la condition 3.9 ou la condition 3.10 n'est pas satisfaite en un nœud m du réseau cœur (cas du sens montant ou du sens descendant), alors il y a congestion. En tout nœud m où apparaît une telle congestion, les sessions $s' \in \mathcal{S}_m$ sont classées par ordre croissant de priorité. Puis tant que les tests 3.9 et 3.10 ne sont pas satisfaits tous les deux, les sessions $s' \in \mathcal{S}_m$ telles que $P(s') < P(s)$ font l'objet, une à une, par ordre croissant de priorité, d'une désallocation paliers par paliers de leur montant de ressources allouées, si cela est possible c'est-à-dire si $\rho_{s'}^{CN} > r_{s'}^{\min}$. Pour une telle session $s' \in \mathcal{S}_m$ la désallocation effectuée est la suivante:

- α est un entier naturel qui détermine le taux de désallocation à chaque palier,
- k est un entier naturel représentant le pas d'incrémentatation correspondant à chaque palier, $0 < k \leq \alpha$,

$$\begin{aligned}\rho_{s'}^{CN} &= r_{s'}^{\max} - k \frac{r_s^{\max} - r_s^{\min}}{\alpha}, \\ \sigma_{s'} &= \sigma_{s'}^k, \\ \Delta_{s'}^m &= \sigma_{s'}^k + L_{s'}^{\max}, \\ \rho_{s'}^{RN} &= \mathcal{R}(r_{s'}^{\max} - k \frac{r_s^{\max} - r_s^{\min}}{\alpha})\end{aligned}$$

avec

$$\sigma_{s'}^k = \left(r_{s'}^{\max} - k \frac{r_s^{\max} - r_s^{\min}}{\alpha} \right) \left(D_{s'}^{q, \max} - \sum_{m=1}^{h_{s'}} \frac{L_{\max}}{r^m} \right) - 2(h_{s'} - 1)L_{s'}^{\max} + L_{s'}^{\max},$$

$D_{s'}^{q, \max}$ étant calculée à partir de l'équation (3.5).

Ainsi les sessions de la classe de priorité la plus faible subissent, une à une, une dégradation uniforme de leur débit alloué jusqu'au niveau acceptable $k = \alpha$ afin de libérer les ressources nécessaires à l'admission de la session entrante. Si le montant des ressources libérées est toujours insuffisant, alors les sessions de la classe de priorité directement supérieure font successivement l'objet de la même réévaluation des ressources allouées.

Finalement, sous condition que les ressources pouvant être libérées soient suffisantes, la session entrante s est acceptée dans le réseau et les ressources des sessions s' concernées sont libérées.

Suite à l'apparition du phénomène de congestion en un nœud m du chemin $\mathcal{P}(s)$ de la session entrante s , les montants des ressources des sessions en cours présentes à ce nœud sont éventuellement réévaluées. Toute réévaluation du montant des ressources allouées au nœud m à la session dégradée s' est répercutée en chacun des nœuds du chemin $\mathcal{P}(s')$ de la session dégradée et sur l'interface radio. Ce contrôle de flot est effectué par la modification des *Rspec* des sessions dégradées, ces états contenant le montant des ressources allouées aux sessions dans le mécanisme RSVP.

3.3 Contrôle d'admission basé sur les mesures

L'architecture de contrôle et de décision décrite dans la section 3.2 permet un contrôle d'admission adaptatif garantissant la qualité de service des usagers dans le réseau. Les descripteurs de trafic $[r_s^{\min}, r_s^{\max}]$ sont des descripteurs à priori du trafic attendu. Cependant, dans le cadre des applications multimédias, il est difficile de savoir à priori quel sera le trafic émis par la source. Nous développons donc une procédure d'admission nommée MBCAC se basant sur des descripteurs de trafic issus de mesures en temps réel, plutôt que sur des descripteurs à priori. Ainsi, en combinant les mesures de trafic avec l'archi-

texture d'admission pour le service garanti décrite à la section 3.2, nous proposons une procédure d'admission qui a pour but de parvenir à un meilleur taux d'utilisation, tout en maintenant une QoS satisfaisante et un taux de blocage (NdS) faible.

3.3.1 Le processus de mesure

3.3.1.1 Implémentation dans le réseau

Le but d'un contrôle d'admission à base de mesures est de transférer la charge de la caractérisation du trafic de la source au réseau afin de réaliser les décisions d'admission non plus sur les spécifications exactes de la source mais sur la base de mesures. La procédure se fonde sur l'utilisation des caractérisations à priori des sources pour les flots demandant l'accès au réseau et des caractérisations issues de mesures pour les flots présents dans le système. Les méthodes de contrôle d'admission établies dans la littérature (section 1.6.3) se basent sur la mesure du trafic agrégé des flots présents dans le système.

Dans le cadre de notre étude, où les flots sont gérés indépendamment les uns des autres, nous considérons que les mesures sont faites individuellement pour chaque session en cours dans le réseau. Ainsi, dans le sens montant, le mécanisme de mesure est implanté au niveau de chaque terminal mobile, afin de caractériser le trafic émis par les sources des sessions en cours. De même, dans le sens descendant, le mécanisme de mesure est implanté au niveau de chaque source de trafic en entrée du réseau cœur. Quelque soit le sens de la session en cours, l'estimation du trafic émis est placée dans la structure de spécification des ressources *RSpec* associée à cette session, ce qui permet de disposer, en chaque nœud du chemin de la session, à la fois du débit en cours alloué et du débit en cours mesuré au niveau de la source.

3.3.1.2 Mécanisme de mesure

La méthode de contrôle d'admission développée dispose de deux composantes distinctes: le processus de mesure qui estime le chargement du réseau et l'algorithme d'admission lui-même qui s'appuie sur cette estimation. Nous choisissons d'utiliser le processus de mesure par fenêtre temporelle utilisé par Jamin *et al.*^[35] et décrite dans la section

1.6.3.

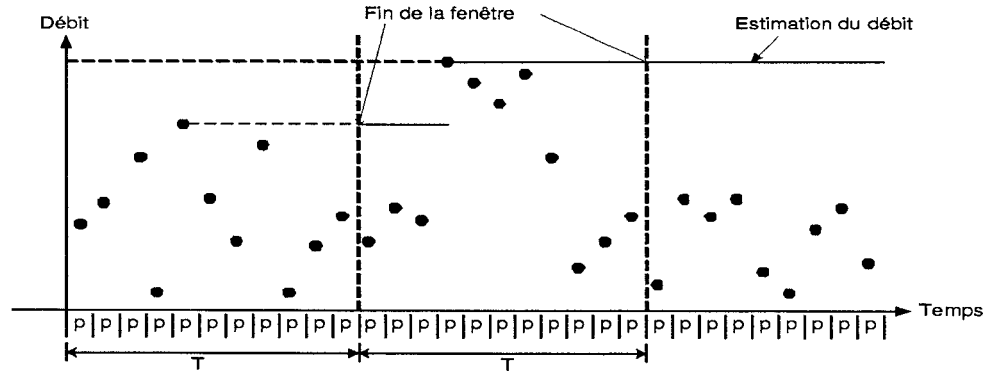


FIG. 3.6 – Mécanisme de mesure

Le processus de mesure utilise les paramètres T et p . La période de moyennage p contrôle la sensibilité de l'estimation: plus p est petit, plus l'estimation est sensible aux bursts de trafic. La taille de la fenêtre T contrôle l'adaptabilité du mécanisme de mesure au chargement du trafic: plus T est petit, plus le mécanisme s'adapte au trafic; cependant, plus T est grand, plus nous obtenons de stabilité dans le mécanisme. Les valeurs données à ces paramètres ont donc un impact conséquent sur les performances de l'algorithme d'admission qui en découle. Ces valeurs doivent être choisies en fonction des caractéristiques du trafic émis et éventuellement en fonction des priorités des sessions dont on mesure le trafic.

Dans notre étude, le trafic envoyé par chaque session en cours dans le réseau est mesuré. On note $\hat{\rho}_s$ l'estimation du débit d'une session s quelconque en cours dans le réseau. La valeur de $\hat{\rho}_s$ est actualisée de la façon suivante:

$$\hat{\rho}_s = \begin{cases} r_s^{\max} & \text{lors de l'admission de la session dans le réseau,} \\ \hat{\rho}_s^p & \text{si } \hat{\rho}_s^p > \hat{\rho}_s, \hat{\rho}_s^p \text{ étant le débit moyen mesuré sur la période } p, \\ \max(\hat{\rho}_s^p) & \text{sur la dernière fenêtre } T. \end{cases}$$

La figure 3.7 décrit l'algorithme de mesure implémenté.

Algorithme de mesure

```

mesure( $\hat{\rho}_s$ )      {lorsqu'une nouvelle fenêtre de mesure débute.}
 $\hat{\rho}_{s,\max}^p = 0$ ;
for  $i = 1$  to  $\frac{T}{p}$  do
     $\hat{\rho}_{s,i}^p =$  débit moyen mesuré sur la  $i$ ème période d'échantillonnage
    if ( $\hat{\rho}_{s,i}^p > \hat{\rho}_{s,\max}^p$ )
        then  $\hat{\rho}_{s,\max}^p = \hat{\rho}_{s,i}^p$ ;   { $\hat{\rho}_{s,\max}^p$  représente la plus grande valeur de débit moyen mesuré.}
    end if
    if  $\hat{\rho}_{s,i}^p > \hat{\rho}_s$                 {la valeur de débit moyen mesurée excède  $\hat{\rho}_s$ , immédiatement
        then  $\hat{\rho}_s = \hat{\rho}_{s,i}^p$ ;        actualiser  $\hat{\rho}_s$  avec la nouvelle valeur moyenne mesurée.}
    end if
end for
 $\hat{\rho}_s = \hat{\rho}_{s,\max}^p$ ; {actualise  $\hat{\rho}_s$  avec la plus grande valeur de débit moyen mesurée sur la
                        fenêtre  $T$ .}

```

FIG. 3.7 – Pseudo-code de l'algorithme de mesure

3.3.2 Spécification de QdS**3.3.2.1 Descripteurs**

La procédure à base de mesures repose sur les mêmes descripteurs de QdS que ceux présentés à la section 3.1.6.1 pour les sessions demandant l'accès au réseau. En revanche, dans cette étude, la session s en entrée du réseau (sens montant ou sens descendant) a une demande en termes de débit égale à $r_s^{\max}$, puis le débit alloué à la session pourra être modifié par la suite en fonction du trafic mesuré, et non plus forcément dans l'intervalle à priori $[r_s^{\min}, r_s^{\max}]$.

Toute session s admise dans le réseau (sens montant ou sens descendant) est entièrement décrite par les descripteurs de trafic présentés à la section 3.1.6.2, auxquels on ajoute l'estimation $\hat{\rho}_s$ du trafic émis par la source initiant la session s .

3.3.2.2 Garantie de QdS

Afin de satisfaire les contraintes de délai et de bande passante exigées par les sessions, nous nous plaçons toujours dans le contexte de garantie établit à la section 3.1.5. Ainsi, dans le contexte basé sur les mesures, une session s entrante se voit allouer un débit $r_s^{\max}$ puis, éventuellement, si les mesures indiquent une sous-utilisation des ressources, la

session en cours peut se voir attribuer une nouvelle bande passante plus faible qui sera alors garantie. La garantie sur le délai, quant à elle, s'opère en ajustant la profondeur du *bucket* en fonction du débit alloué à la session et du délai demandé.

Cependant, comme pour tout service à base de mesures, la précision et l'incertitude sur les mesures ont un impact important sur les performances de l'algorithme d'admission. De plus, le fait de faire des mesures a des conséquences sur la QoS. En effet, en caractérisant le trafic des sources à l'aide de mesures, on fait le pari que le passé récent est un bon indicateur pour l'avenir proche. Si ce pari n'est pas tenu, alors la QoS offerte est dégradée. Dans le cadre de notre étude, par exemple, une session en cours qui sous-utilise ses ressources (d'après l'estimation établie par l'algorithme de mesure) peut se voir attribuer une bande passante plus petite que celle actuellement allouée, tout en garantissant la borne sur le délai exigée. Cependant, si dans l'avenir le trafic émis par la source avant le lissage devient supérieur à la bande passante allouée à la session en cours, alors la source n'est plus conforme aux paramètres alloués à la session dans le réseau et la QoS en termes de débit n'est plus satisfaite. Ainsi, l'algorithme d'admission proposé est développé dans un contexte de service garanti et la QoS offerte est acceptable tant que le processus de mesure est un bon indicateur du futur.

3.3.3 Vérification des ressources disponibles

Les notations utilisées pour la description de l'algorithme MBCAC sont exactement celles présentées à la section 3.2.1.

Toute session s demande à être admise dans le réseau au débit de transmission $r_s^{\max}$. La procédure d'admission de la session est celle présentée à la section 3.2.2. Les tests d'admission respectivement sur l'interface radio et sur les nœuds du réseau cœur sont ceux établis et détaillés dans les sections 3.2.2.1 et 3.2.2.2.

3.3.4 Réserve et allocation des ressources

Ainsi, comme dans le cas de la procédure décrite en 3.2.4, toute session s admise dans le réseau, que ce soit dans le sens montant ou dans le sens descendant se voit attribuer les ressources suivantes lors de son admission:

$$\begin{aligned}
\rho_s^{\text{CN}} &:= r_s^{\text{max}}, \\
\sigma_s &:= \sigma_s^{\text{max}}, \\
\Delta_s^m &:= \sigma_s^{\text{max}} + L_s^{\text{max}} \quad \forall m \in \{1, \dots, h_s - 1\}, \\
\rho_s^{\text{RN}} &:= \mathcal{R}(r_s^{\text{max}}), \\
\text{FER}_s &:= F_s
\end{aligned}$$

avec

$$\sigma_s^{\text{max}} = r_s^{\text{max}} \left(D_s^{q, \text{max}} - \sum_{m=1}^{h_s} \frac{L_{\text{max}}}{r^m} \right) - 2(h_s - 1)L_s^{\text{max}} + L_s^{\text{max}},$$

$D_s^{q, \text{max}}$ étant calculée à partir de l'équation (3.5).

3.3.5 Libération de ressources

Nous nous plaçons dans le cas où l'un au moins des tests d'admission 3.7 (si sens montant), 3.8 (si sens descendant), 3.9 et 3.10 a échoué. Dans ce cas nous sommes dans un état de congestion, et il n'y a pas suffisamment de ressources dans le réseau pour accepter la session s qui se présente en entrée. Alors les ressources des sessions en cours dans le réseau, de plus faible priorité que s et dont le trafic mesuré indique une sous-utilisation des ressources, peuvent éventuellement être libérées afin d'admettre la session entrante s de plus haute priorité. A priori il ne s'agit pas d'une renégociation de QoS étant donné que l'on désalloue des ressources non-utilisées. Cependant nous avons vu que cette désallocation entraîne le risque d'une possible dégradation de QoS en termes de débit pour l'avenir; c'est pourquoi il s'agit de libérer des ressources disponibles uniquement chez les sessions en cours de plus faible priorité.

3.3.5.1 Congestion sur l'interface radio

Si la condition 3.7 (si sens montant) ou la condition 3.8 (si sens descendant) n'est pas satisfaite, alors il y a congestion sur l'interface radio. Les sessions $s' \in \mathcal{S}$ sont alors classées dans l'ordre de priorité croissante. Puis, tant que le test d'admission sur la partie radio

n'est pas satisfait, les sessions $s' \in \mathcal{S}$ telles que $P_{s'} < P_s$ font l'objet, une à une, par ordre croissant de priorité, d'une désallocation de leur montant de ressources allouées si les mesures indiquent une sous-utilisation des ressources, c'est-à-dire si $\rho_{s'}^{\text{RN}} - \mathcal{R}(\hat{\rho}_{s'} + \epsilon) > 0$. Pour une telle session $s' \in \mathcal{S}$ la désallocation effectuée est la suivante:

$$\begin{aligned}\rho_{s'}^{\text{RN}} &:= \mathcal{R}(\hat{\rho}_{s'} + \epsilon), \\ \rho_{s'}^{\text{CN}} &:= \hat{\rho}_{s'} + \epsilon, \\ \sigma_{s'} &:= \hat{\sigma}_{s'}, \\ \Delta_{s'}^m &:= \hat{\sigma}_{s'} + L_{s'}^{\text{max}}\end{aligned}$$

avec

$$\hat{\sigma}_{s'} = (\hat{\rho}_{s'} + \epsilon) \left(D_{s'}^{q, \text{max}} - \sum_{m=1}^{h_{s'}} \frac{L_{\text{max}}}{r^m} \right) - 2(h_{s'} - 1)L_{s'}^{\text{max}} + L_{s'}^{\text{max}},$$

$D_{s'}^{q, \text{max}}$ étant calculée à partir de l'équation (3.5).

Le paramètre ϵ est une valeur très inférieure à $\hat{\rho}_{s'}$ permettant de lever une certaine incertitude sur la mesure faite, et garder une marge de sécurité par rapport à la précision de la mesure.

De plus, il apparaît que $\rho_{s'}^{\text{CN}}$ et $\sigma_{s'}$ peuvent prendre des valeurs très faibles si les mesures indiquent au moment de la congestion une très faible activité du trafic émis par s' . Or si on affaiblit trop la valeur allouée à $\sigma_{s'}$, alors pour continuer à garantir le délai il faudrait que la source soit quasiment à taux constant puisque le *bucket* n'absorberait presque plus les bursts de trafic, aussi petits soient ils. De ce fait, nous imposons une autre condition pour autoriser la libération de ressources inutilisées chez une session s' en cours: $\hat{\rho}_{s'} > \underline{\rho}_{s'}$, où $\underline{\rho}_{s'}$ est une borne inférieure sur le débit que l'on peut allouer à une session s' .

Ainsi, sous condition que les ressources pouvant être libérées soient suffisantes, la session entrante s est acceptée dans le réseau et les ressources des sessions s' concernées sont libérées.

Lors de l'admission d'une session s , les montants de ressources allouées aux sessions s' en cours dans la même cellule que s font éventuellement l'objet d'une réévaluation. Cette

réévaluation du montant des ressources alloué à une session s' en cours est répercutée en chacun des nœuds du chemin $\mathcal{P}(s')$ sur le réseau cœur en mettant à jour les ressources allouées dans la structure $RSpec$ du mécanisme RSVP. Le contrôle de flot est effectué par la modification des $Rspec$ des sessions dégradées.

3.3.5.2 Congestion en un nœud du réseau cœur

Si la condition 3.9 ou la condition 3.10 n'est pas satisfaite en un nœud m du réseau cœur (cas du sens montant ou du sens descendant), alors il y a congestion. En tout nœud m où apparaît une telle congestion, les sessions $s' \in \mathcal{S}_m$ sont classées par ordre croissant de priorité. Puis tant que les tests 3.9 et 3.10 ne sont pas satisfaits tous les deux, les sessions $s' \in \mathcal{S}_m$ telles que $P(s') < P(s)$ font l'objet, une à une, par ordre croissant de priorité, d'une désallocation de leur montant de ressources allouées, si le trafic mesuré indique une sous-utilisation des ressources c'est-à-dire si $\rho_{s'}^{\text{CN}} - \hat{\rho}_{s'} > \epsilon$. Il faut également vérifier que $\hat{\rho}_{s'} > \underline{\rho}_{s'}$, où $\underline{\rho}_{s'}$ est la borne inférieure sur le débit que l'on peut allouer à une session s' comme définie à la section 3.3.5.1. Pour une telle session $s' \in \mathcal{S}_m$ la désallocation effectuée est la suivante:

$$\begin{aligned}\rho_{s'}^{\text{CN}} &= \hat{\rho}_{s'} + \epsilon, \\ \sigma_{s'} &= \hat{\sigma}_{s'}, \\ \Delta_{s'}^m &= \hat{\sigma}_{s'} + L_{s'}^{\text{max}}, \\ \rho_{s'}^{\text{RN}} &= \mathcal{R}(\hat{\rho}_{s'} + \epsilon)\end{aligned}$$

avec

$$\hat{\sigma}_{s'} = (\hat{\rho}_{s'} + \epsilon) \left(D_{s'}^{q, \text{max}} - \sum_{m=1}^{h_{s'}} \frac{L_{s'}^{\text{max}}}{r^m} \right) - 2(h_{s'} - 1)L_{s'}^{\text{max}} + L_{s'}^{\text{max}},$$

$D_{s'}^{q, \text{max}}$ étant calculée à partir de l'équation (3.5).

Ainsi, sous condition que les ressources pouvant être libérées soient suffisantes, la session entrante s est acceptée dans le réseau et les ressources des sessions s' concernées sont libérées.

Suite à l'apparition du phénomène de congestion en un nœud m du chemin $\mathcal{P}(s)$ de la

session entrante s , les montants des ressources des sessions en cours présentes à ce nœud sont éventuellement réévaluées. Toute réévaluation du montant des ressources allouées au nœud m à la session s' est répercutée en chacun des nœuds du chemin $\mathcal{P}(s')$ de la session dégradée et sur l'interface radio. Ce contrôle de flot est effectué par la modification des *Rspec* des sessions dégradées, ces états contenant le montant des ressources allouées aux sessions dans le mécanisme RSVP.

3.3.6 Algorithme d'admission MBCAC

Les figures 3.8 et 3.9 présentent les pseudo-codes de l'algorithme à base de mesures, dans le sens montant et dans le sens descendant. On rappelle que l'admission doit se faire sur le réseau radio et le réseau cœur.

3.4 Conclusion

En conclusion, l'algorithme d'admission traite les flots de manière indépendante, cela permettant d'assurer une garantie sur le délai pour chaque session. Etant donné que la bande passante sur la partie radio est beaucoup moins importante que sur les réseaux fixes hauts débits, il semble tout à fait possible de déployer un tel algorithme d'admission par flots plus complexe qu'un algorithme traitant des flots agrégés. D'autre part, un algorithme de mesure est implémenté pour augmenter l'utilisation de la bande passante en estimant le trafic de chaque flot. Le contrôle d'admission intégrant les mesures continue de garantir la borne sur le délai mais fournit un service prédictif pour la garantie sur le débit. Des pertes peuvent également apparaître dans ce contexte, mais la nature même du lien radio peu fiable permet d'envisager le contrôle d'admission à base de mesures comme une méthode intéressante. L'algorithme d'admission développé, avec ou sans les mesures, est adaptatif en réallouant les ressources des sessions déjà actives dans le réseau en fonction de la congestion à l'admission et en fonction de la nature du trafic entrant. Le chapitre suivant teste les performances de l'algorithme à base de mesures, comparativement à celui basé sur des descripteurs de trafic à priori.

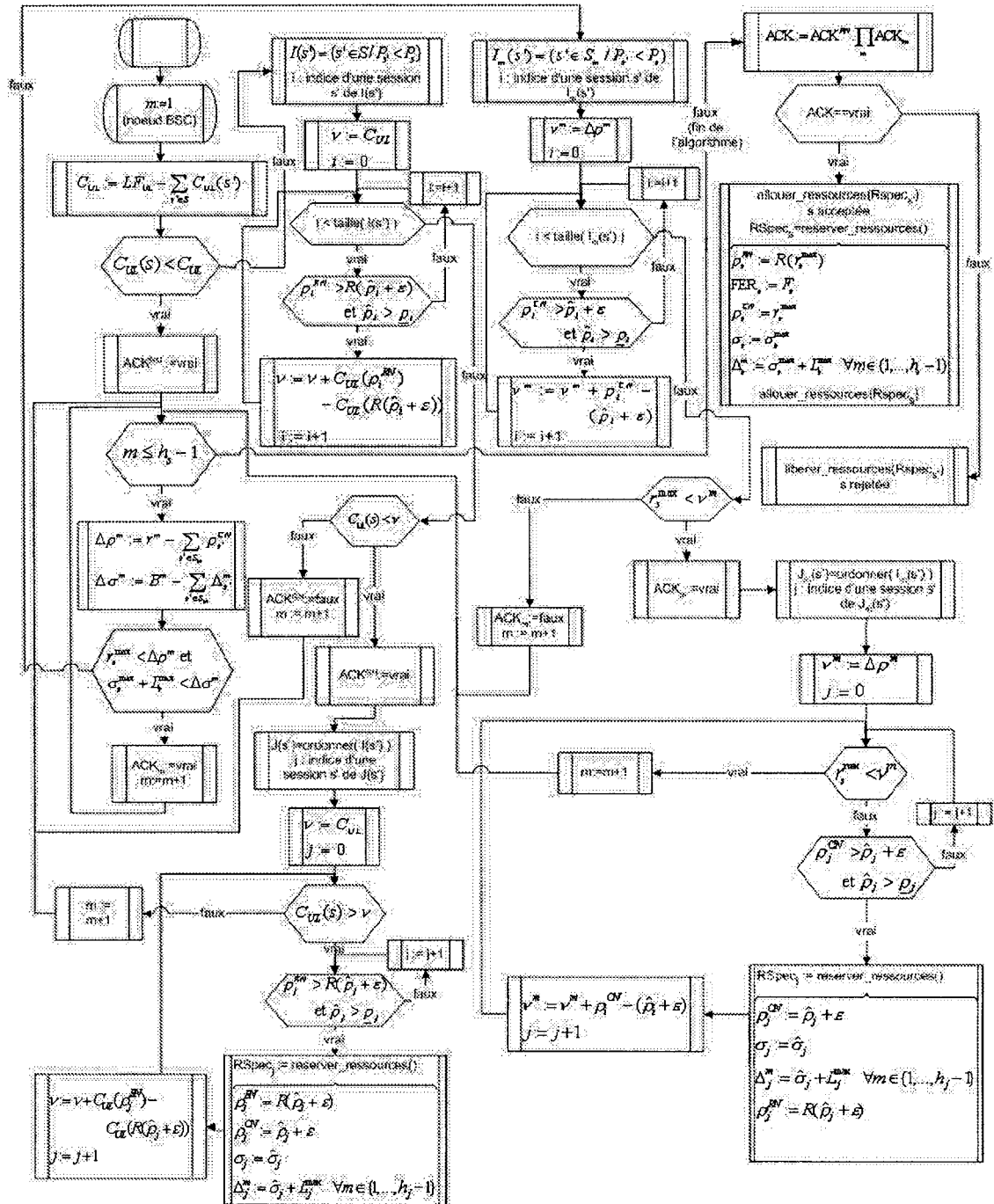


FIG. 3.8 – Algorithme MBCAC sens montant

CHAPITRE 4

RÉSULTATS

L'objet de notre étude consiste à proposer une structure de contrôle et de décision pour la gestion des ressources dans les futurs réseaux de communications mobiles et d'évaluer les performances et l'influence des paramètres choisis dans le cadre de la conception du contrôle d'admission. La démarche adoptée met l'accent sur la gestion des ressources limitées dans le réseau afin d'offrir des garanties de QoS aux applications multimédias à contraintes différenciées. L'étude a porté sur l'établissement d'équations de capacité radio pour la partie radio et sur le choix de la discipline d'ordonnancement WFQ pour la partie réseau cœur ainsi que sur la conception de contrôles d'admission bout en bout adaptatifs présentés au chapitre 3. Les algorithmes CAC0 et CAC1 ont été étudiés dans le réseau cœur par Troeung^[7]. Dans notre étude ces algorithmes ont été repris et adaptés pour établir des contrôles d'admission bout en bout sur le réseau d'accès 3G. De plus, un algorithme à base de mesures a été implémenté à fins de comparaison des performances avec l'algorithme CAC1.

Notre étude en termes de résultats portera donc sur les performances et les caractéristiques des algorithmes sur l'interface radio. En effet, le lien radio est la partie congestionnante pour les futurs réseaux cellulaires: les ressources radio étant beaucoup plus limitées que les ressources du réseau cœur, le challenge pour ces réseaux est essentiellement la gestion et le dimensionnement de celles-ci. Nous allons donc analyser les performances d'une cellule radio. Les objectifs de simulation sont les suivants:

- analyser les performances de l'algorithme MBCAC, en termes d'utilisation des ressources radio, de taux de blocage (NdS), et de garantie de QoS, en fonction des paramètres de mesures choisis,
- comparer les performances des algorithmes MBCAC et CAC1,
- étudier une approche possible pour le dimensionnement du réseau d'accès radio.

4.1 Description du simulateur

Le simulateur a été développé sous le logiciel de modélisation de réseaux OPNET^[46]. Les algorithmes d'admission et les politiques associées ont été développés dans cet environnement tout en assurant l'intégration avec les modules déjà existant dans le logiciel. En particulier, le protocole de réservation RSVP^[10] tel qu'implanté dans OPNET est utilisé comme protocole de support pour la signalisation du contrôle de flot évoqué à la section 3.1.6.2.

4.1.1 Réseau simulé

Dans le cadre des résultats présentés, le réseau simulé est celui de la figure 4.1.

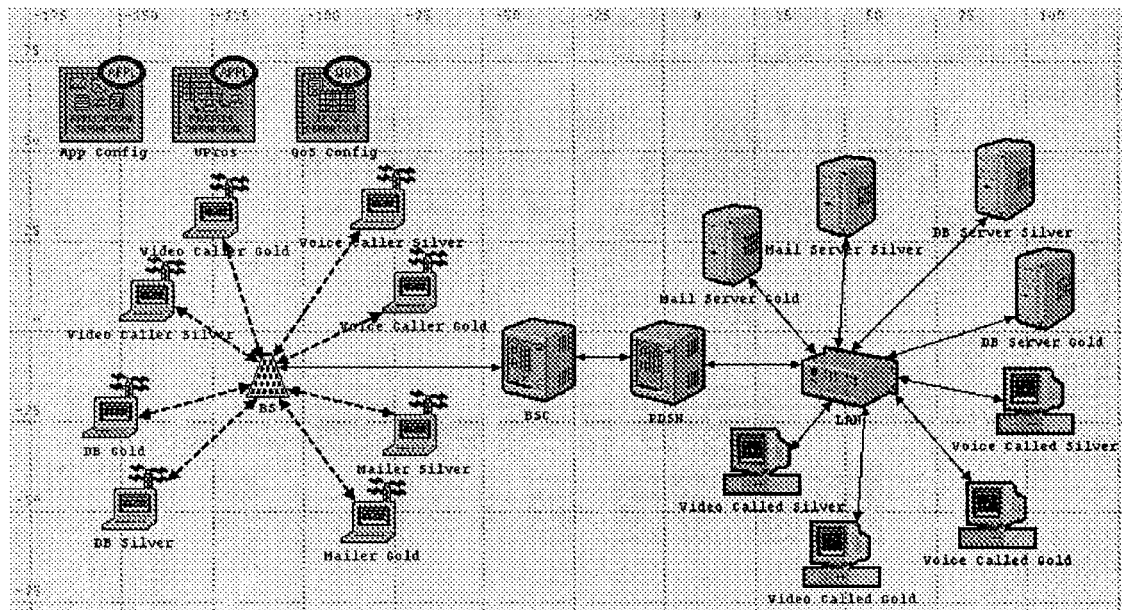


FIG. 4.1 – Topologie de l'environnement simulé

Les liens radios sont modélisés entre les terminaux mobiles et la station de base BS, c'est-à-dire que le délai radio ainsi que les débits discrets de la configuration radio choisie y sont implantés. On rappelle que l'étude porte sur les performances de l'accès radio, étant interconnexion avec le réseau cœur. D'après l'homogénéité du trafic supposée dans les équations de capacité radio, chaque cellule radio se comporte identiquement à

ses voisines. C'est pourquoi le trafic dans une seule cellule est simulé. Le lien entre la station de base et le nœud BSC du réseau cœur a une capacité supposée infinie car nous considérons qu'il n'y a pas de congestion sur ce lien.

Le réseau cœur, très simplifié pour notre étude, est constitué du nœud BSC et du nœud PDSN qui assure l'interconnexion du réseau cœur avec le réseau externe. L'interconnexion entre les nœuds BSC et PDSN est réalisée par un lien duplex de capacité limitée. Cependant, nous choisissons pour nos simulations une capacité de haut débit afin que la congestion du trafic d'une cellule radio ait lieu sur le réseau radio.

Les classes de trafic présentées à la section 1.4.2 sont mises en œuvre par des applications types. Une répartition d'applications types est soumise en entrée du réseau (réseau cœur pour le sens descendant, et réseau radio pour le sens montant). De plus seuls les niveaux de priorités *Gold* et *Silver* sont considérés.

TAB. 4.1 – *Sessions d'utilisateurs*

Classe	Application	Description	Hierarchie
<i>Tâche de fond</i>	Mail Silver	envoi et réception de courriels	0
<i>Tâche de fond</i>	Mail Gold	envoi et réception de courriels	1
<i>Interactive</i>	DB Silver	requêtes de base de données	2
<i>Interactive</i>	DB Gold	requêtes de base de données	3
<i>Diffusion</i>	Video Silver	vidéo <i>streaming</i>	4
<i>Diffusion</i>	Video Gold	vidéo <i>streaming</i>	5
<i>Conversationnelle</i>	Voice Silver	VoIP	6
<i>Conversationnelle</i>	Voice Gold	VoIP	7

Comme l'indique la table 4.1, nous considérons 8 types de sessions ordonnées suivant les priorités des usagers et des applications. Dans le cadre de notre étude, toute session est initiée par un des terminaux mobiles du réseau d'accès radio et est bilatérale: elle comporte un flot montant et un flot descendant de même priorité. On considère donc que la session bilatérale est composée d'une session dans le sens montant et une autre dans le sens descendant comme précisé dans la section 3.2.1. Etant donné que les trafics multimédias susceptibles d'être offerts sur les futurs systèmes de communications mobiles ont une charge de trafic asymétrique avec une charge plus importante sur le lien descendant, les ressources demandées par la session montante diffèrent de celles demandées par

la session descendante. La session bilatérale est admise dans le réseau si les ressources sont disponibles dans le sens montant et dans le sens descendant.

4.1.2 Modèles de trafic

4.1.2.1 Caractéristiques des applications

Les futurs réseaux 3G transporteront à la fois des applications non temps réel et des applications temps réel. Les applications temps réel (voix et vidéo) ont la particularité d'être codées et compressées à l'émission avant d'être encapsulées dans les paquets IP. Il existe plusieurs types de codeurs de qualité et de caractéristiques différentes pour la voix et également pour la vidéo. Les codeurs standards pour l'audio recommandés par l'UIT sont présentés dans la table 4.2, et les codeurs standards principaux utilisés actuellement pour la vidéo sont présentés dans la table 4.3. Pour la voix, les trames de données audio sont compressées puis transmises au réseau suivant un débit constant. La vidéo, elle, est la succession d'une série d'images de taille fixée envoyées à des intervalles de temps réguliers. Le système de compression compresses chaque image puis ces dernières sont transmises sur le réseau. Les codeurs permettent ainsi de réduire la bande passante utilisée en réduisant l'espace mémoire (en bits) utilisé par chaque image. Cependant, en conservant une qualité donnée, chaque image est plus ou moins compressible. Ainsi il existe deux types de codeurs: les codeurs à taux fixe remplissent simplement un espace mémoire qui est ensuite écoulé à un débit constant sur le réseau. Il s'agit alors, étant donné le débit choisi, de compresser chaque image de manière à ce que la mémoire ne soit jamais vide et ne déborde jamais. Cela implique éventuellement une réduction de la qualité des images qui ne peuvent être assez bien compressées. En revanche, les codeurs à taux variable choisissent une qualité d'image donnée, et étant donnée cette qualité ils compressent chaque image au maximum, et il en résulte un trafic à débit variable. Les codeurs à taux variables sont appelés à être plus utilisés dans le futur car ils ne dégradent pas la qualité de l'image.

Pour les futurs terminaux multimédias, les codeurs principalement utilisés seront le codeur H.263 ou MPEG4 à taux variable pour la vidéo et le codeur G.729 à 8kbps pour la

TAB. 4.2 – *Standards audios recommandés par l'UIT*

Standard	Débit (kbps)	Taille d'une trame (ms)	Année
G.711 - audio codec	64	0.125	1972
G.726, G.727 - audio codec	16, 24, 32, 40	0.125	1990
G.722 - audio codec	48, 56, 64	0.125	1988
G.728 - speech codec	16	0.625	1992/1994
G.729 - speech codec	8	10	1995
G.723 - speech codec	5.3 et 6.4	30	1995

TAB. 4.3 – *Standards vidéos principaux*

Standard	Taux de compression	Application
H.261	p×64kbps (p= 1 to 30)	vidéo conférence
H.263	<64kbps à plus de 2Mbps	vidéo conférence
MPEG-1	de 1.0 à 1.5 Mbps	stockage d'images animées sur des systèmes de sauvegarde (CD-ROM)
MPEG-2	de 4 à 20 Mbps dépendant de la qualité	télévision digitale
MPEG-4	9.6 à 1024 kbps	applications multimédia (vidéo conférence)

voix. Ainsi, les applications voix générées dans notre simulateur utilisent le codeur G.729 disponible dans OPNET, et les applications de vidéo à la demande sont générées avec un débit variable par la donnée d'une taille d'image (en bits) qui suit une loi non constante et d'un temps d'interarrivée des images qui est constant. Toutes les applications types mises en œuvre sont des instances des modèles d'applications proposés par OPNET Technologies Inc. Les paramètres nécessaires à la configuration des instances d'applications sont définis dans les manuels de configuration fournis par OPNET Technologies Inc^[47].

La table 4.4 définit les paramètres descriptifs des applications temps réel et non temps réel mises en œuvre sur le lien montant et sur le lien descendant dans le simulateur. Pour une même application, seul le niveau de priorité (*Gold* ou *Silver*) peut changer. Les paramètres pour la voix sont ceux pris par Opnet par défaut avec le codeur G.729; pour les autres applications nous nous sommes inspirés des valeurs données par Troeung^[7] et avons parfois augmenté la taille des messages ou requêtes afin de modéliser du trafic multimédia.

La table 4.5 décrit les exigences de QoS propres à chaque application. Nous les avons choisies de manière à respecter les débits générés par la configuration de chaque applica-

TAB. 4.4 – Paramètres de configuration des applications types dans le simulateur

		sens montant	sens descendant
VoIP	Longueur du silence (en s)	exponentiel(0.65)	exponentiel(0.65)
	Longueur de la conversation (en s)	exponentiel(0.352)	exponentiel(0.352)
	Type de codeur	G.729	
	Nombre de trames par paquets	1	
Vidéo	Temps d'interarrivée des images (en s)	constant(1)	constant(0.15)
	Tailles des images (en bytes)	exponentiel(60)	exponentiel(1000)
DB	Requêtes-Total transactions	100%	
	Temps d'interarrivée des transactions (en s)	exponential(15)	
	Taille des requêtes (en kbytes)	constant(130)	
E-mail	Temps d'interarrivées (en s)	exponentiel(10)	exponentiel(8)
	Taille des messages (en kbytes)	constant(60)	constant(75)

tion donnée dans la table 4.4.

TAB. 4.5 – Paramètres des exigences de QoS des applications dans le simulateur

		$r_s^{\min}$ (en bps)	$r_s^{\max}$ (en bps)	$D_s^{\max} + \delta D_s$ (en ms)	F_s
E-mail	sens montant	28000	72000	6000	5%
	sens descendant	32000	120000	6000	5%
DB	sens montant	28000	72000	10000	5%
	sens descendant	72000	120000	10000	5%
Vidéo	sens montant	8000	8000	1500	5%
	sens descendant	72000	120000	1500	5%
Voix	sens montant	8000	8000	300	5%
	sens descendant	8000	8000	300	5%

4.1.2.2 Profils de trafic

Chaque terminal mobile modélisé sur la figure 4.1 est une source de trafic générant un certain type de session (application et priorité). Le trafic généré par chaque terminal est la succession en série du même type de profil de trafic défini statistiquement comme décrit par Troeung^[7] et détaillé ci-dessous.

Pour notre étude, nous utilisons un nombre variable de terminaux mobiles de type *Voice Gold*, *Voice Silver*, *Video Gold*, *Video Silver*, *DB Gold*, *DB Silver*, *Mail Gold* et *Mail Silver*.

La figure 4.2 indique la succession en série du même profil défini statistiquement pour

un terminal donné. Les paramètres descriptifs de la succession de profils sont les suivants:

- début de profil : distribution statistique,
- durée : distribution statistique,
- temps d'inter-répétition (intervalles entre les arrivées) : distribution statistique,
- nombre de répétitions : constant ou illimité,
- mode de répétition : série.

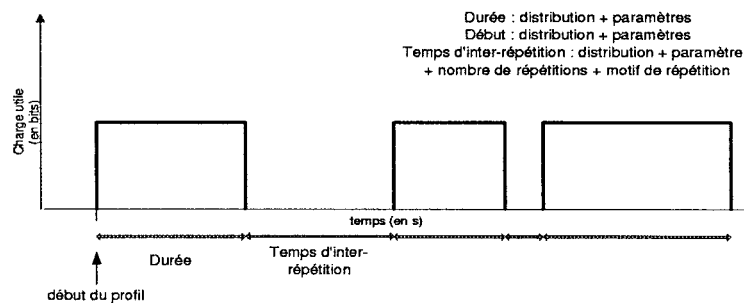


FIG. 4.2 – *Succession temporelle des profils de trafic*

Le profil de trafic associé à un terminal consiste en la succession de sessions en parallèle présentant les mêmes caractéristiques de trafic, d'exigences de QoS et de niveau de priorité.

La figure 4.3 indique la succession des applications dans un profil de trafic appartenant à un terminal donné. Les paramètres descriptifs de la succession d'applications sont les suivants :

- nombre de sessions : constant,
- décalage de début de session : distribution statistique,
- durée : distribution statistique,
- répétitivité : unique après le début du profil de trafic.

À l'intérieur d'un même profil de trafic, les applications se succèdent donc sur le mode parallèle. Dans notre simulateur, nous générons une et une seule session à l'intérieur d'un profil. Chaque application (ou session) est instanciée d'après le modèle d'application

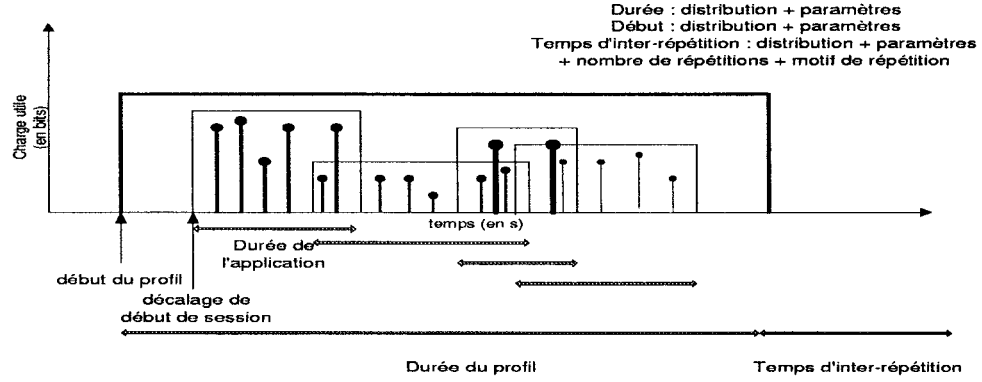


FIG. 4.3 – Succession temporelle des sessions pour un profil de trafic considéré

original avec les valeurs données dans le tableau 4.4 afin de produire la quantité de trafic attendue.

Ainsi, pour notre étude tous les trafics générés par les terminaux sont des successions d'applications en série. D'autre part les différents scénarios que nous allons considérer seront déterminés par le nombre de terminaux utilisés et le profil de trafic donné à chaque terminal mobile.

4.1.3 Paramètres de configuration annexes

Du fait de la capacité élevée du lien sur le réseau cœur, on suppose que le délai de propagation D_s^p et le délai de transmission D_s^{tr} pour toute session s sont supposés négligeables par rapport au temps d'attente dans les files d'attente du réseau cœur. Quant au temps de traitement D_s^t , on le prend égal à une borne supérieure de $100ms$. L'équation (3.5) devient donc $D_s^{q,max} = D_s^{max} + \delta D_s - D_s^t - \overline{D}_{trame} \times \frac{L_s^{max}}{T_{trame} \rho_s^{RN}}$.

Les paramètres de configuration du réseau cœur constitué des nœuds BSC et PDSN et d'un lien duplex sont décrits dans la table 4.6. En ce qui concerne la taille maximale d'un paquet, sous IP elle est de 1500 octets, et cela comprend les 40 octets d'entêtes nécessaires à l'encapsulation. Cependant, la voix et la vidéo avec les standards utilisés voient la taille maximale de leur paquet pouvant être respectivement de 50 octets et de 296 octets, comme cela est précisé par ElGebaly^[48]. Ces données permettent le calcul du

délai sur le lien radio comme spécifié par la formule (3.2).

TAB. 4.6 – Paramètres pour le réseau cœur

Capacité du lien	DS3 (44 Mbps)	
Taille de paquet maximale	VoIP	50 octets
	Vidéo <i>streaming</i>	296 octets
	Courriels	1500 octets
	Requêtes bases de données	1500 octets

Les paramètres de configuration pour l'interface radio sont récapitulés dans la table 4.7. Ce sont ceux utilisés à la section 2.4.1 lors de l'analyse des équations de capacité radio.

TAB. 4.7 – Configuration des paramètres radios

		Paramètre	Valeur		
		W	1.25×10^6 Hz		
		N_0	1.2589×10^{-20} W/Hz		
		P_{BS}	20 W		
Paramètre	Valeur	F_{CONT}	0.2	Paramètre	Valeur
W	1.25×10^6 Hz	w	0.4	T_{trame}	20ms
λ	0.55	LF_{DL}	0.6		
SHG_{UL}	0.54	SHG_{DL}	0.54		
LF_{UL}	0.6	ν_s^{DL}	voix: 0.67		
ν_s^{UL}	voix: 0.67		autres: 1		
	autres: 1	K	3600		
p_{SH}^{UL}	p_{SH}^{DL}	R	1000 m		
		d_0	1500 m		
		Distribution	Suivant la distance		
(a)		(b)		(c)	
sens montant		sens descendant		paramètres annexes	

Etant données les valeurs de configuration radio ci-dessus mentionnées et les demandes en termes de débit et de FER décrites dans la table 4.5, nous pouvons établir les résultats présentés dans la table 4.8 qui permettront d'analyser les résultats de simulation par la suite. La notion de capacité dans le sens montant et dans le sens descendant est expliquée au chapitre 2.

TAB. 4.8 – *Capacité utilisée par chaque type de session*

		$\mathcal{R}(r_{\min})$ (bps)	Capacité utilisée	Capacité max
DL	Mail	38400	0.74	8.0
	DB	76800	1.52	
	Video	76800	1.52	
	Voice	9600	0.12	
		$\mathcal{R}(r_{\max})$ (bps)	Capacité utilisée	Capacité max
DL	Mail	153600	3.03	8.0
	DB	153600	3.03	
	Video	153600	3.03	
	Voice	9600	0.12	

		$\mathcal{R}(r_{\min})$ (bps)	Capacité utilisée	Capacité max
UL	Mail	38400	0.041	0.6
	DB	38400	0.041	
	Video	9600	0.010	
	Voice	9600	0.007	
		$\mathcal{R}(r_{\max})$ (bps)	Capacité utilisée	Capacité max
UL	Mail	76800	0.083	0.6
	DB	76800	0.083	
	Video	9600	0.010	
	Voice	9600	0.007	

4.2 Les différents cas d'étude

Les différents scénarios choisis pour étudier les performances de l'algorithme d'admission sont caractérisés par une certaine répartition du trafic selon le type d'application. Ces scénarios doivent donc être conçus de manière à représenter les profils de demandes et la nature des chargements de trafic pour les futurs réseaux de troisième génération. Ils se basent sur des pronostics, à savoir la façon dont les usagers utiliseront les applications. Les deux scénarios proposés s'inspirent des répartitions données dans le rapport technique de Lind^[49] rédigé chez Ericsson, et qui ont pour but d'étudier les caractéristiques des nouvelles applications multimédias. Ces deux scénarios types sont caractérisés par la contribution en termes de bande passante de chaque classe de trafic, qui est évaluée en pourcentage du trafic total. Le premier scénario (A) prédit ainsi une large utilisation des applications multimédias, tandis que le second scénario (B) privilégie la voix tout en considérant une augmentation des applications multimédias par rapport à l'utilisation actuelle des services de données.

4.2.1 Scénario (A)

Ce scénario test est destiné à l'évaluation du déploiement des nouvelles applications plutôt qu'à l'évaluation de l'application de la parole (VoIP). Le trafic généré par les sources en entrée du réseau ainsi que leur demande en bande passante maximum et minimum sont présentés dans la figure 4.4.

Le tableau 4.9 présente les différents profils de trafic générés. Les valeurs ont été fixées afin de générer la quantité et la répartition de trafic désirées. Notons qu'à un instant quelconque, il y a au plus 20 sessions simultanées dans la cellule radio (8 sessions voix, 4 sessions vidéo, 4 sessions mail et 4 sessions DB). Etant donnés les résultats de la table 4.8 sur les équations de capacité radio, cela peut déjà représenter une saturation de la capacité d'une cellule pour du trafic multimédia.

Sur le temps de simulation, cette quantité de trafic a représenté 111 sessions ayant fait une requête au réseau. La figure 4.5 montre la répartition moyenne du trafic généré en entrée du réseau sur la période de simulation.

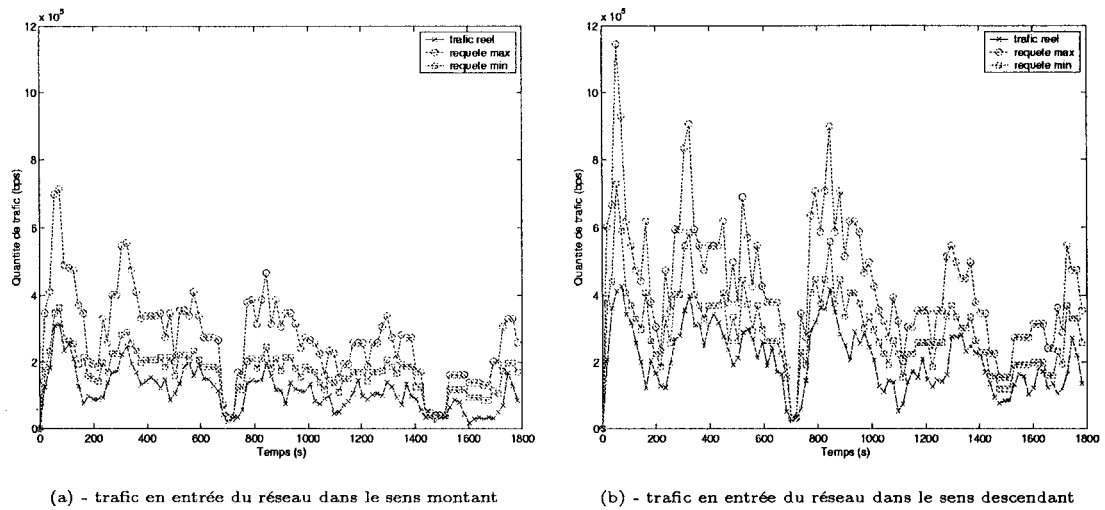
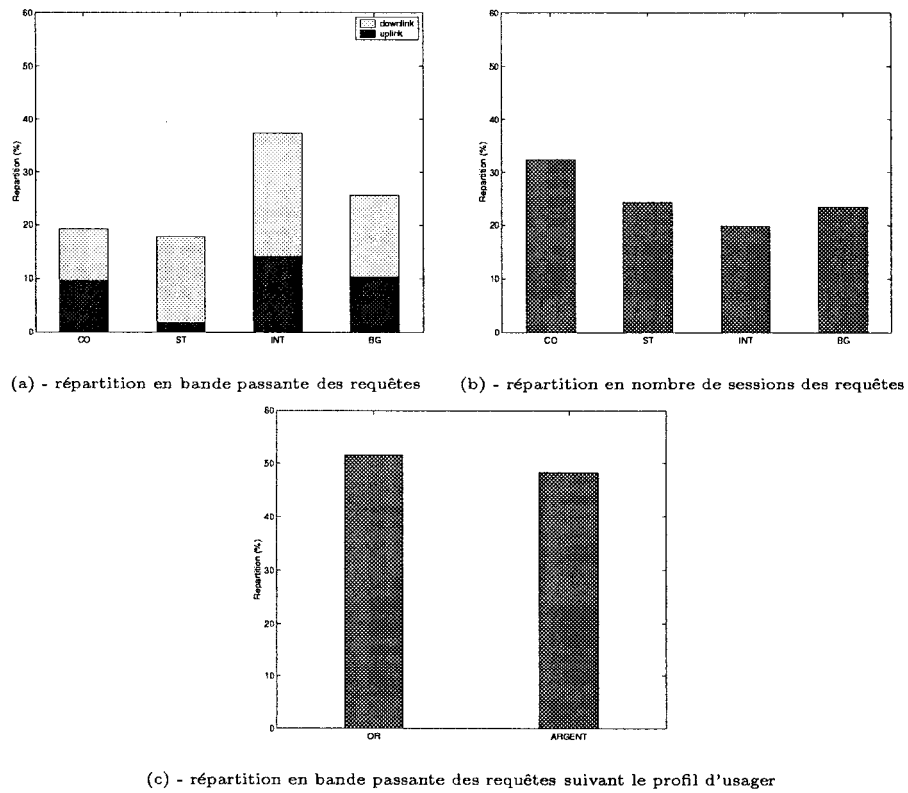
TAB. 4.9 – *Paramètres de configuration des profils de trafic pour le scénario (A)*

Terminal	Début profil et session (s)	Durée profil (s)	Durée session (s)	Inter-répétition du profil (s)
Mail Silver ¹	constant(15)	exponentiel(125)	exponentiel(60)	constant(155)
Mail Gold ¹	constant(20)	exponentiel(125)	exponentiel(60)	constant(155)
DB Silver ¹	constant(25)	exponentiel(400)	exponentiel(150)	constant(50)
DB Gold ¹	constant(30)	exponentiel(400)	exponentiel(150)	constant(50)
Video Silver ¹	constant(35)	exponentiel(200)	exponentiel(100)	constant(60)
Video Gold ¹	constant(40)	exponentiel(200)	exponentiel(100)	constant(60)
Voice Silver ^{1,2,3,4}	constant(45)	exponentiel(175)	exponentiel(150)	constant(5)
Voice Gold ^{1,2,3,4}	constant(50)	exponentiel(175)	exponentiel(150)	constant(5)
Mail Silver ²	constant(55)	exponentiel(125)	exponentiel(60)	constant(155)
Mail Gold ²	constant(60)	exponentiel(125)	exponentiel(60)	constant(155)
DB Silver ²	constant(65)	exponentiel(400)	exponentiel(150)	constant(50)
DB Gold ²	constant(70)	exponentiel(400)	exponentiel(150)	constant(50)
Video Silver ²	constant(75)	exponentiel(200)	exponentiel(100)	constant(60)
Video Gold ²	constant(80)	exponentiel(200)	exponentiel(100)	constant(60)

4.2.2 Scénario (B)

Ce scénario test tient compte du déploiement des applications nouvelles ainsi que des applications de parole (VoIP). Le trafic généré par les sources en entrée du réseau ainsi que leur demande en bande passante maximum et minimum sont présentés dans la figure 4.6.

Le tableau 4.10 présente les profils de trafic générés dans ce scénario. Les valeurs ont été fixées afin de générer la quantité et la répartition de trafic désirées. A un instant quelconque, on voit qu'il y a au plus 25 sessions simultanées actives dans la cellule radio

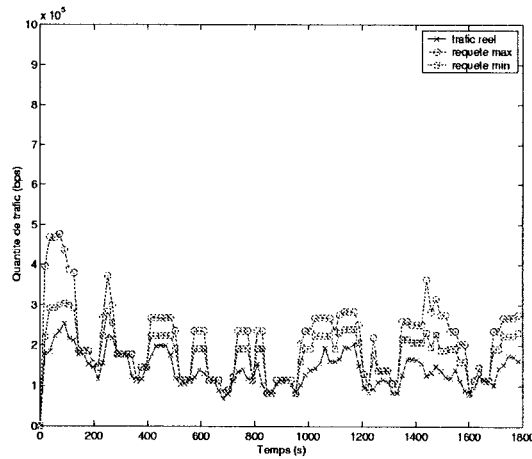
FIG. 4.4 – *Quantité de trafic pour le scénario (A)*FIG. 4.5 – *Profil de trafic pour le scénario (A)*

(16 sessions voix, 4 sessions vidéo, 2 sessions mail et 3 sessions DB).

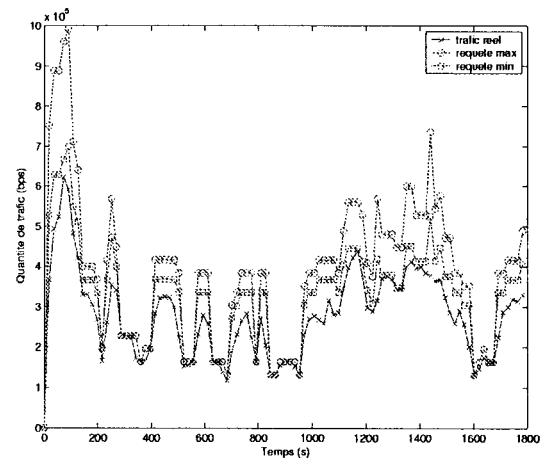
Sur le temps de simulation, cette quantité de trafic a représenté 119 sessions ayant fait une requête au réseau. La figure 4.7 montre la répartition moyenne du trafic généré en entrée du réseau sur la période de simulation.

TAB. 4.10 – Paramètres de configuration des profils de trafic pour le scénario (B)

Profil de trafic	Début profil et session (s)	Durée profil (s)	Durée session (s)	Inter-répétition du profil (s)
Mail Silver ¹	constant(15)	exponentiel(250)	exponentiel(60)	constant(155)
Mail Gold ¹	constant(20)	exponentiel(250)	exponentiel(60)	constant(155)
DB Silver ¹	constant(25)	exponentiel(400)	exponentiel(150)	constant(200)
DB Gold ¹	constant(30)	exponentiel(400)	exponentiel(150)	constant(200)
Video Silver ¹	constant(35)	exponentiel(500)	exponentiel(100)	constant(300)
Video Gold ¹	constant(40)	exponentiel(500)	exponentiel(100)	constant(300)
Voice Silver ^{1,2,3,4}	constant(45)	exponentiel(350)	exponentiel(300)	constant(50)
Voice Gold ^{1,2,3,4}	constant(50)	exponentiel(350)	exponentiel(300)	constant(50)
DB Gold ²	constant(70)	exponentiel(400)	exponentiel(150)	constant(200)
Video Silver ²	constant(75)	exponentiel(500)	exponentiel(100)	constant(300)
Video Gold ²	constant(80)	exponentiel(500)	exponentiel(100)	constant(300)
Voice Silver ^{5,6,7,8}	constant(85)	exponentiel(350)	exponentiel(300)	constant(50)
Voice Gold ^{5,6,7,8}	constant(90)	exponentiel(350)	exponentiel(300)	constant(50)



(a) - trafic en entrée du réseau dans le sens montant



(b) - trafic en entrée du réseau dans le sens descendant

FIG. 4.6 – Quantité de trafic pour le scénario (B)

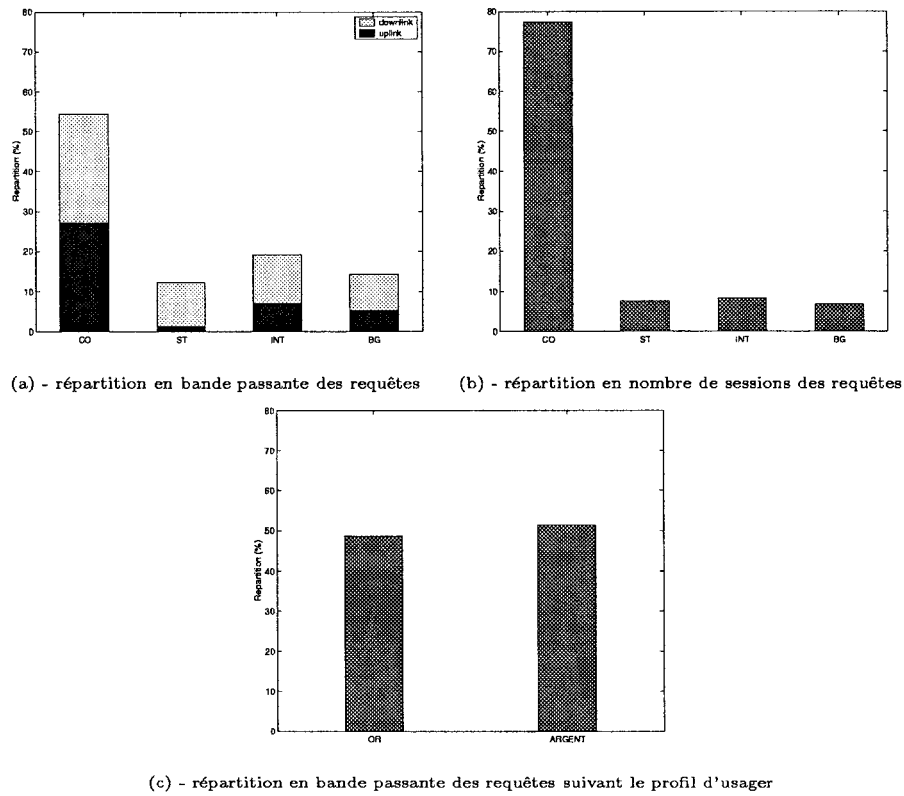
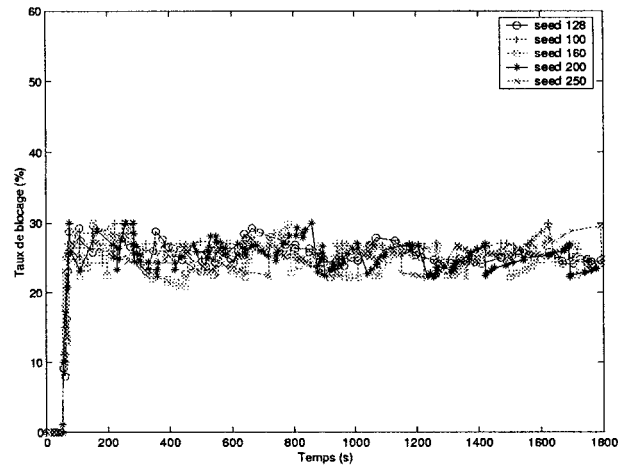


FIG. 4.7 – Profil de trafic pour le scénario (B)

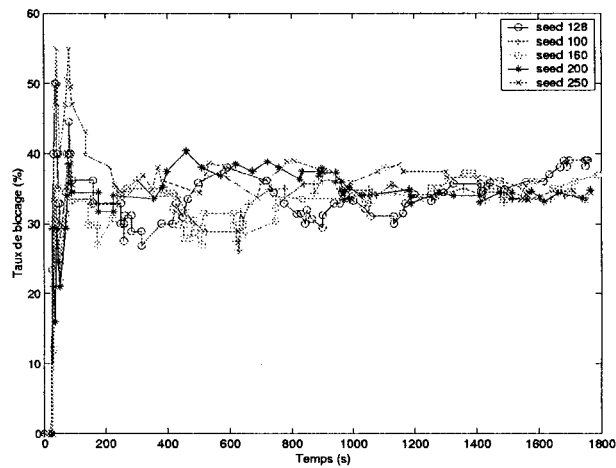
4.3 Validation de performances

Dans notre étude, les performances et caractéristiques des algorithmes d'admission sont obtenues par le biais de la simulation. Dans cette section nous présentons sur deux exemples, la variabilité des résultats obtenus en fonction du germe original qui permet de générer aléatoirement le trafic dans OPNET. En effet, le trafic est généré sur la base de critères probabilistes afin de modéliser le comportement réel des sources et applications, et donc pour un même scénario, les résultats ne sont pas exactement les mêmes lorsque la racine initiale des séquences aléatoires est modifiée. Nous avons effectué 5 simulations avec 5 germes différents. Ainsi la figure 4.8 montre la faible variabilité des résultats en fonction de la valeur du germe initial (*seed*). Il est donc raisonnable par la suite de déterminer les performances du système sur la base d'un seul et unique germe initial. De plus, l'évolution des tendances, la comparaison des performances et les caractéristiques

des procédures de contrôle d'admission peuvent être totalement déduites des résultats obtenus sur la base d'un seul germe.



(a) - taux de blocage pour le scénario A sous la procédure CAC1



(b) - taux de blocage pour le scénario B sous la procédure MBCAC

FIG. 4.8 – Taux de blocage globaux obtenus pour 5 séquences de simulations selon le scénario A avec la procédure CAC1 et selon le scénario B avec la procédure MBCAC

4.4 Etudes et performances de l'algorithme MBCAC en fonction des paramètres de mesure

L'algorithme utilisé pour mesurer le trafic émis par chaque source se base sur le mécanisme de fenêtre développé par Jamin *et al.*^[35]. Ils mettent en évidence que la taille de la fenêtre de moyennage p et la taille de la fenêtre de mesure T ont un impact important sur les performances du système. Dans le cadre de notre étude, dans un contexte radio, nous évaluons également l'influence de ces paramètres de mesure sur différents critères de performance: le taux de blocage global, le taux d'occupation à la station de base et la garantie de QoS. Cela permettra ensuite de se placer dans une configuration optimale du MBCAC et de comparer ses performances avec celles de l'algorithme CAC1. Comme indiqué par Jamin *et al.*^[35], nous choisissons le paramètre p égal environ au temps de transmission de 100 paquets et le paramètre T égal à plusieurs fois la valeur de p .

4.4.1 Phénomène de blocage

Nous évaluons pour chaque scénario le pourcentage de sessions refusées en fonction premièrement de la taille de la fenêtre T , puis en fonction de la longueur de l'échantillonnage p . Notons que les tailles des fenêtres sont exprimées en équivalent en nombre de paquets transmis.

A travers l'un et l'autre des scénarios représentés sur la figure 4.9, il apparaît que plus la taille de la fenêtre T est grande, p étant fixé, plus le taux de rejet est important. En effet, la valeur de débit mesurée est d'autant plus souvent actualisée que la taille de la fenêtre T est petite. Plus T est petit, plus il y a de chances que le débit mesuré soit adapté au trafic et qu'éventuellement il indique une sous utilisation des ressources. En revanche si T devient grand, l'actualisation de la valeur du débit mesurée est moins fréquente et moins adaptée au trafic, et il en résulte alors un taux de blocage supérieur.

Il est à noter que l'influence de la taille T sur le taux de blocage dépend du type de trafic émis. En effet, pour le scénario B, l'impact de la taille de cette fenêtre est beaucoup moins net que dans le cas du scénario A. En effet, le scénario B est chargé à 50% par de la voix, qui a un débit constant; ainsi sur le trafic voix, l'algorithme de mesure n'apporte

pas d'avantages, et les bénéfices en termes de taux de blocage engendrés par les mesures sur les autres types de trafic sont moins marqués que dans le cas du scénario A qui est fortement chargé en trafic à taux variable.

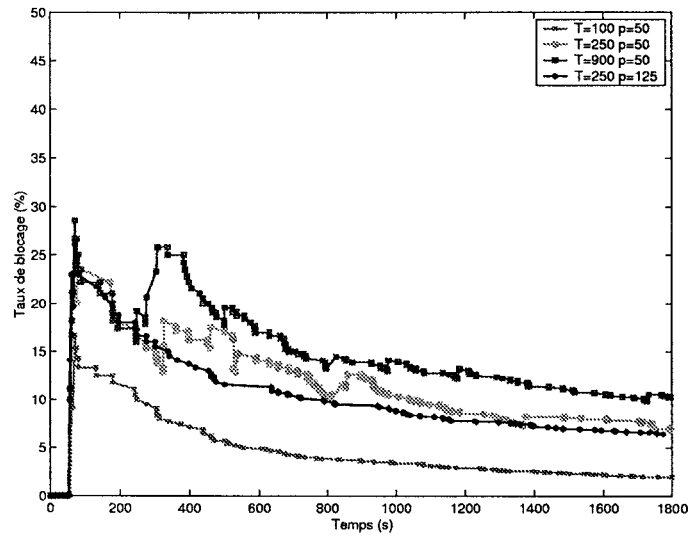
En ce qui concerne l'influence du paramètre p , on s'aperçoit que le taux de blocage diminue légèrement lorsque p croît, T étant fixé. En effet, lorsque p devient grand le moyennage lisse mieux le débit mesuré et détecte moins les bursts de trafic; ainsi, lors de l'actualisation de la valeur de débit à la fin de la fenêtre T , les échantillons de débits moyens calculés sur les intervalles de temps p sont plus lissés et ont des valeurs plus faibles que pour une taille de p plus petite qui permettrait de détecter des bursts. Comme l'on actualise la valeur du débit mesurée par la moyenne la plus élevée sur la fenêtre T qui vient de passer, cette valeur est d'autant plus faible que le paramètre p est grand, et cela engendre plus de probabilité de pouvoir libérer des ressources et donc d'obtenir un taux de blocage plus faible.

Cependant la sensibilité du taux de blocage au paramètre p n'est pas très forte étant donné une valeur de T donnée. Ainsi pour chaque scénario, nous obtenons une faible variation du taux de blocage pour p valant 50 ou 125 paquets. En fait, pour p autour de 100 paquets, les variations sur les performances sont très faibles, c'est pourquoi nous montrons simplement le cas où $p = 50$ et $p = 125$ qui sont les valeurs extrêmes à donner à p pour rester autour de 100 paquets.

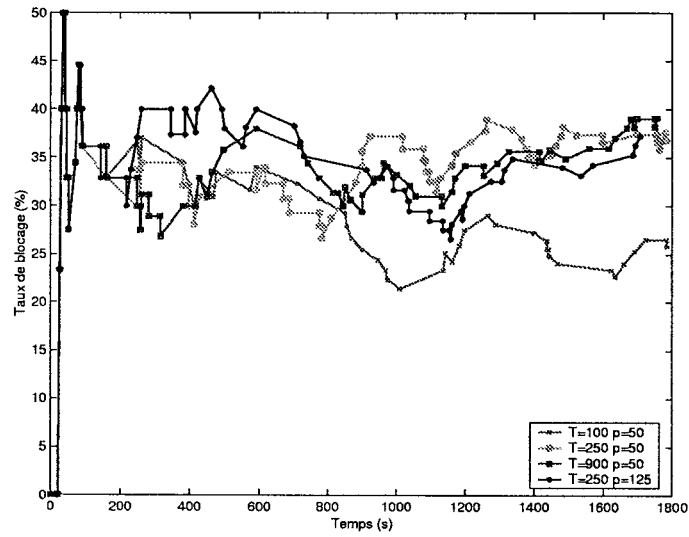
Ainsi nous retenons qu'en termes de taux de blocage, il est avantageux de choisir une taille de fenêtre T la plus petite possible étant donnée la taille de p . De plus, même si la sensibilité du taux de blocage au paramètre p est assez faible, le gain en termes de taux de blocage est meilleur lorsque p augmente (autour de 100 paquets).

4.4.2 Taux d'occupation à la station de base

La figure 4.10 indique les taux d'occupation moyens à la station de base en termes de capacité radio consommée, sur le lien montant et sur le lien descendant, pour chacun des scénarios, et en fonction de paramètres de mesure. L'observation des taux d'utilisation de la capacité sur les liens montant et descendant indique bien la prépondérance du lien descendant dans le déploiement des applications multimédias.



(a) - scénario A: taux de blocage global en fonction de T et p

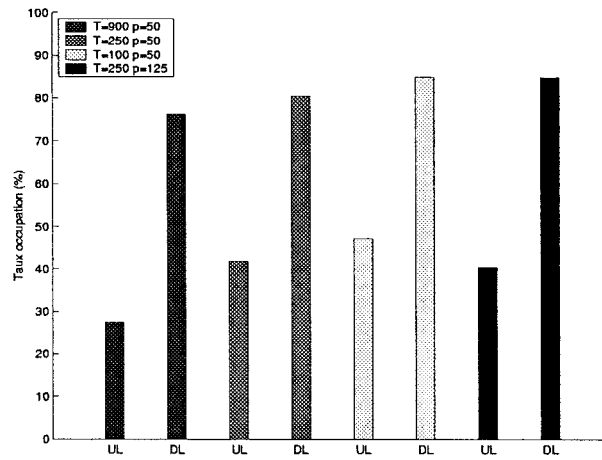


(b) - scénario B: taux de blocage global en fonction de T et p

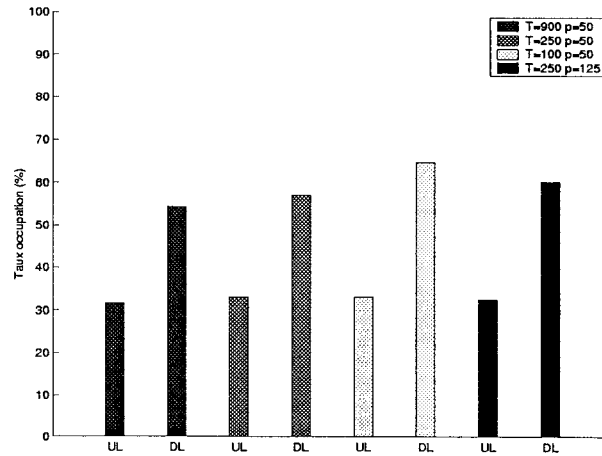
FIG. 4.9 – NdS globaux en fonction des paramètres de mesure pour les scénarios A et B

De même que le taux de blocage est d'autant plus faible que la valeur de T est petite, le taux d'utilisation de la ressource radio est d'autant plus important que T est petit. En effet, plus T prend une petite valeur, plus l'algorithme est capable de répartir un montant plus élevé de ressources parmi un plus grand nombre de sessions.

D'autre part, on voit que le paramètre p pris entre les valeurs 50 et 125 n'influence quasiment pas le taux d'utilisation des ressources à la station de base. On note toutefois une légère amélioration du taux d'utilisation lorsque p passe de 50 à 125.



(a) - scénario A: taux d'occupation à la BS en fonction de T et p



(b) - scénario B: taux d'occupation à la BS en fonction de T et p

FIG. 4.10 – Taux d'utilisation de la capacité disponible à la BS en fonction des paramètres de mesure pour les scénarios A et B

TAB. 4.11 – *Taux d'utilisation comparés pour les scénarios A et B en fonction des paramètres de mesure*

	scénario A					
	utilisation UL			utilisation DL		
	moy (%)	min (%)	max (%)	moy (%)	min (%)	max (%)
T=900 p=50	27.4	6.5	76.1	76.2	23.6	89.9
T=250 p=50	41.9	8.8	76.1	80.5	29.1	97.9
T=100 p=50	47.1	7.1	79.1	84.9	36.6	97.9
T=250 p=125	40.4	4.7	79.2	84.7	32.8	97.9

	scénario B					
	utilisation UL			utilisation DL		
	moy (%)	min (%)	max (%)	moy (%)	min (%)	max (%)
T=900 p=50	31.7	13.5	64.1	54.3	15.7	94.3
T=250 p=50	33.2	10	67.1	57.0	18.9	97.5
T=100 p=50	33.2	16.5	67.1	64.7	15.2	97.5
T=250 p=125	32.6	10	67.1	60.2	18.9	99.3

4.4.3 Garantie de QdS

La procédure MBCAC libère une partie des ressources allouées aux sessions en cours lorsque le mécanisme de mesure détecte une sous-utilisation, et ce dans but de pouvoir accommoder au mieux les sessions entrantes. Les débits de transmission affectés aux sessions en cours subissent donc des fluctuations, le niveau de QdS en termes de débit n'est à priori pas affecté puisque le débit alloué est dégradé si le mécanisme de mesure indique que ce nouveau débit suffira à satisfaire la demande en débit de la session en cours. En ce qui concerne le niveau de QdS en termes de délai, il s'agit de vérifier que les contraintes de délai sont bien respectées tout au long de la durée de vie des sessions.

La figure 4.11 indique les fonctions de densité cumulées du délai bout en bout pour les applications de voix et de base de données. L'observation des courbes montre d'abord que la contrainte de délai pour chaque application est strictement respectée. D'autre part les variations sur le délai sont relativement faibles. En effet, le respect de la contrainte de délai est assuré par le délai dans les files d'attente des nœuds du réseau cœur que nous contrôlons au travers du dimensionnement adéquat des *token buckets* et des mémoires tampons pour chaque session. Ainsi, le délai moyen pour une application voix est de 255 ms et le délai moyen de l'application DB est de 1.33 s, et ce quels que soient les scénarios et les valeurs des paramètres de mesure. Avec la technique d'ordonnancement WFQ, les

flots sont isolés les uns des autres, le délai d'un paquet dans la file d'attente d'une session ne varie qu'en fonction du comportement de sa source et ne varie pas en fonction des autres flots. Ainsi les faibles variations sur le délai observées pour chaque cas sont dues au comportement de chaque source. Etant donnée l'échelle des temps considérée, on peut dire que quelle que soit la taille des fenêtres de mesure, le délai est respecté pour les applications temps réel comme pour les applications non temps réel, et qu'il n'y a pas d'impact de la taille des fenêtres sur la valeur de ce délai.

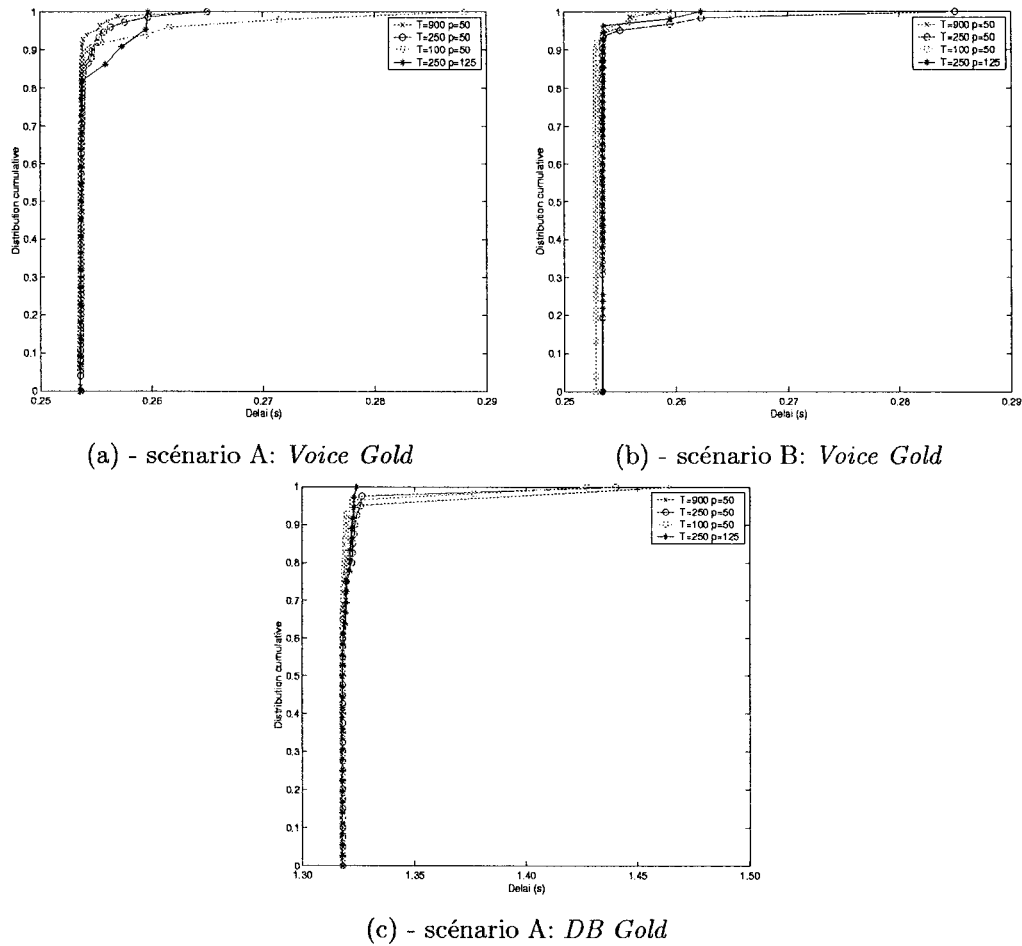


FIG. 4.11 – Distributions cumulatives de la répartition du délai bout en bout pour les applications Voice Gold et DB Gold en fonction des paramètres de mesure

La QoS en termes de délai est donc entièrement satisfaite, ce qui est totalement garanti par la borne de Parekh et Gallager^[16]. En ce qui concerne le débit, nous avons

déjà dit que la procédure MBCAC diminue la bande passante allouée à une session en cours si le mécanisme de mesure estime que la source envoie effectivement ses paquets à un débit inférieur à celui alloué. Ainsi, si le mécanisme de mesure était parfait et était un très bon indicateur du comportement de la source, alors le fait de diminuer la bande passante allouée n'engendrerait pas de dégradation de QdS. Or il se peut que, suivant les valeurs des paramètres de mesure, le mécanisme de mesure sous-évalue le besoin réel de la source en termes de débit. Si l'on est dans ce cas et que les ressources sont libérées à ce moment là, alors nous dégradons la QdS offerte à la session en question. Rappelons que dans ce cas, il doit également se produire des pertes de paquets en avant du *token bucket* de la source, le temps que cette dernière s'adapte au nouveau lissage qui lui est imposé.

La figure 4.12 présente le trafic émis (et accepté dans le réseau sous le scénario A) par un terminal de sessions DB ainsi que les valeurs données par l'algorithme de mesure suivant les paramètres de mesure. On remarque que pour p fixé à 50, plus la fenêtre T est grande moins les mesures sont adaptées au trafic. Cependant, on voit également que si T devient relativement petit, alors les mesures deviennent un plus mauvais indicateur du futur, car certains points sous évaluent les besoins réels de la source en termes de débit. Par ailleurs, pour T fixé à 250, on voit que lorsque p vaut 125, le moyennage lisse plus le trafic que pour $p = 50$ et cela entraîne également certains points de mesure à se trouver en dessous du trafic réellement émis; ainsi la session pourrait voir son débit alloué prendre cette valeur, et alors par la suite cela engendrerait une dégradation de QdS pour la session concernée. Les paramètres de mesure doivent donc être fixés en fonction de la nature du trafic envoyé. Dans le cas des applications générées par le logiciel OPNET, comme cela est le cas pour l'application DB dont le trafic mesuré est représenté ici, il est judicieux de prendre la configuration où $p = 50$ et $T = 250$. En effet, dans cette configuration, et étant donnée la nature du trafic généré, nous assurons que le mécanisme de mesure est un bon indicateur du comportement de la source, et ne sous évalue jamais les ressources nécessaires. En même temps, ces valeurs permettent aux mesures d'être suffisamment bien adaptées au trafic pour demeurer un excellent descripteur du trafic émis. Dans notre étude, la QdS en termes de débit est donc quasiment toujours respectée.

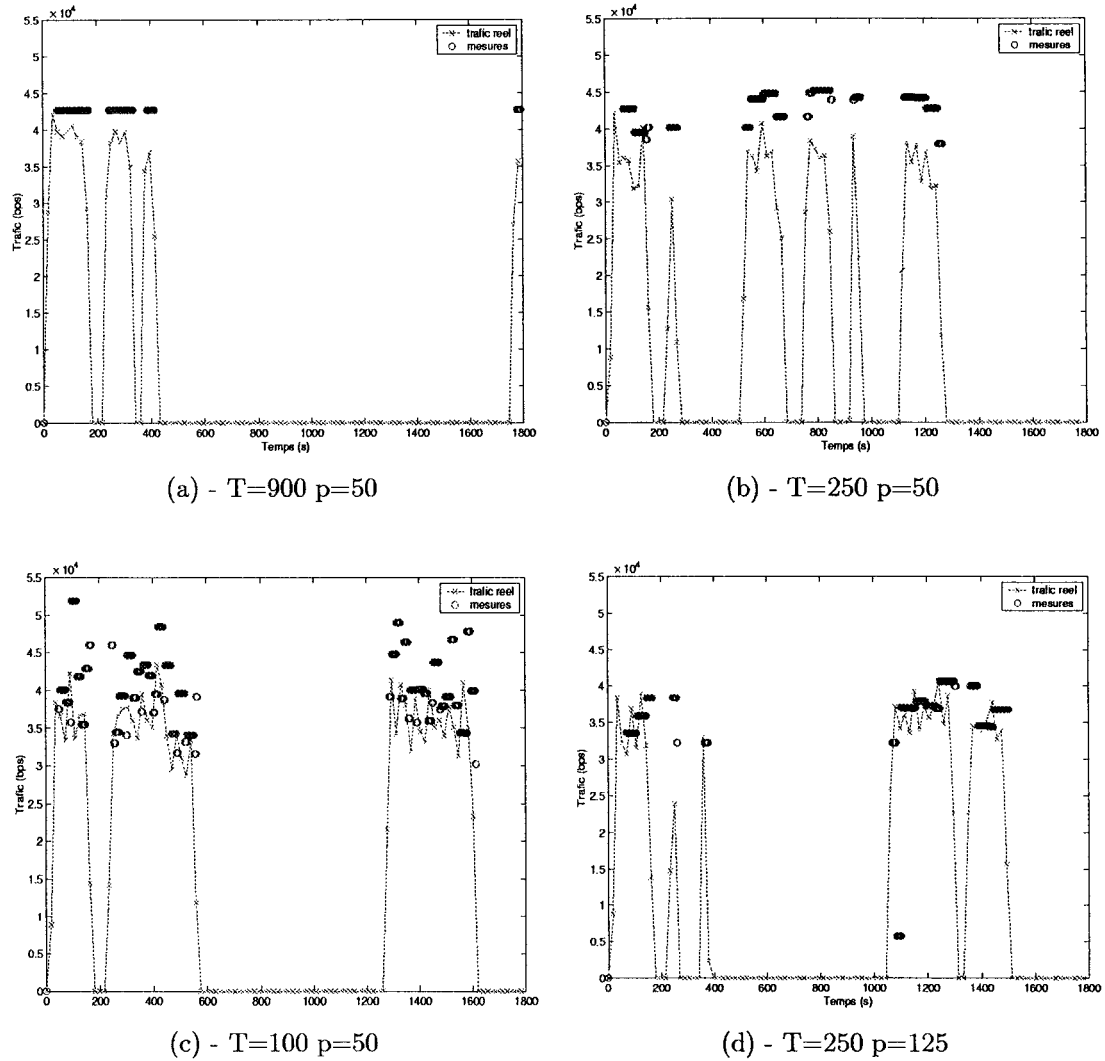


FIG. 4.12 – Mesures du trafic DB dans le sens descendant en fonction des paramètres de mesure

4.5 Etude comparative avec la procédure CAC1

Dans cette section, nous réalisons l'évaluation des performances de l'algorithme de contrôle d'admission MBCAC par comparaison aux performances obtenues avec l'algorithme CAC1.

4.5.1 Phénomène de blocage

Le déploiement de la procédure MBCAC doit permettre de réaliser un contrôle d'admission adaptatif et sélectif, en se basant sur des descripteurs de trafic issus des mesures, tandis que la procédure CAC1 se base sur des descripteurs à priori. Dans les deux contrôles d'admission, une fraction des ressources déjà allouées sont libérées par la dégradation du débit attribué aux sessions de faible priorité afin de pouvoir accommoder les sessions entrantes qui seraient autrement rejetées. Le MBCAC doit, par le biais des mesures, permettre de donner de meilleures performances en termes de taux de blocage et de taux d'utilisation des ressources tout en assurant la garantie de QdS.

Les figures 4.13 et 4.14 présentent les taux de blocage comparés entre les procédures MBCAC et CAC1 pour le scénario A et pour le scénario B. Pour chaque scénario, on observe la présence d'une phase de transition correspondant à la mise en place des sessions dans un système initialement vide et au repos, puis l'instauration d'une phase permanente correspondant à l'arrivée des sessions dans un système possédant une charge en cours.

Les cas d'études A et B correspondent à une charge de trafic comparable, avec un profil régulier au long de la période d'activité du système. Cependant, le scénario A est proportionnellement plus chargé en trafic multimédia sur toute la période d'activité, tandis que le scénario B est chargé à plus de 50% (en bande passante utilisée sur le période d'activité) par de la voix.

D'après le tableau 4.8, on voit que le sens descendant (le sens congestionnant pour les applications multimédias) est saturé autour de 400kbps. Ainsi le tableau 4.12 récapitule quelques données intéressantes.

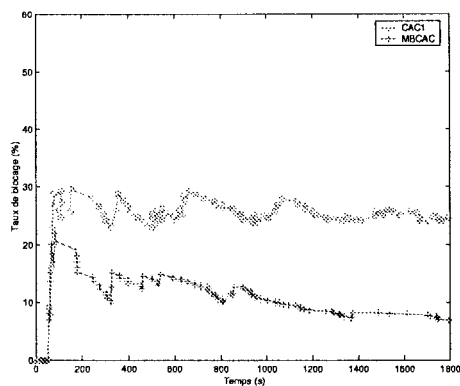
TAB. 4.12 – *Capacité du système dans le sens descendant*

	Quantité sessions Mail	Quantité sessions DB	Quantité sessions Video	Quantité sessions Voice	Capacité utilisée	Bande passante utilisée (kbps)
$\mathcal{R}(r_{\max})$	1	1	0	16	7.98	460.8
	0	0	2	16	7.98	460.8
$\mathcal{R}(r_{\min})$	2	2	2	3	7.92	412.8
	1	2	2	9	7.9	432
	0	0	4	16	8.0	460.8

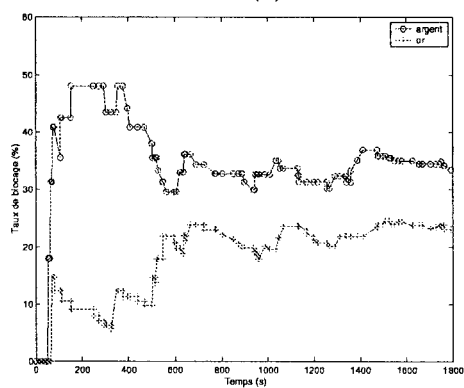
Les courbes (a) des figures 4.13 et 4.14 montrent les taux de blocage globaux sous chaque procédure. Pour chacun des scénarios, la procédure MBCAC est plus capable de libérer les ressources que la procédure CAC1 et le taux de blocage s'en trouve réduit. Ainsi sur la période de stabilité, le taux de blocage moyen est de 8% pour le MBCAC et de 25% pour le CAC1 sous le scénario A, de 35% pour le MBCAC et de 50% pour le CAC1 sous le scénario B. Dans le scénario B, l'accumulation fréquente d'un plus grand nombre de sessions voix dans le système rend les ressources libérables plus rares du fait de leur niveau élevé de priorité. Il est donc plus difficile pour les deux procédures de libérer des ressources quand cela est nécessaire, et il en résulte un taux de blocage plus important. Notons que les taux de blocage sont dus au blocage des sessions dans le sens descendant qui est rapidement saturé en présence de trafic multimédia. Etant donnés les trafics générés, un taux de blocage de 8% sous la procédure MBCAC pour le scénario A par exemple, signifie que sur les 111 sessions ayant fait une requête au réseau durant la période de simulation, 103 ont pu être admises. Les fortes différences de taux de blocage entre les procédures ou les différents scénarios proviennent du fait que les rejets de sessions ou les diminutions de débit se font sur la base des débits discrets disponibles sur le lien radio, qui sont peu nombreux et assez éloignés les uns des autres.

Les courbes (b) des figures 4.13 et 4.14 montrent les taux de blocage suivant les types d'utilisateurs. Pour chacun des scénarios, le trafic en entrée du réseau est constitué d'environ la même quantité de sessions *Gold* que de sessions *Silver*. Il en résulte que pour chaque procédure, le taux de blocage des sessions *Gold* et *Silver* sont équivalents sur les périodes de stabilité. On note toutefois que le blocage des sessions *Silver* est légèrement supérieur à celui des sessions *Gold*.

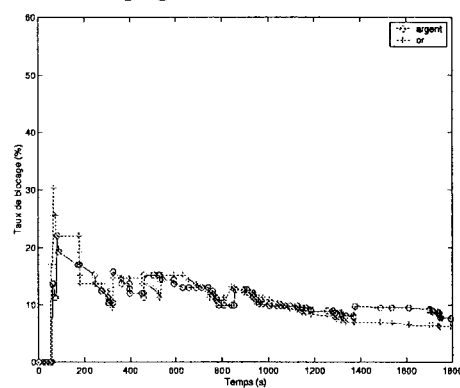
Finalement, les courbes (c) des figures 4.13 et 4.14 montrent les taux de blocage suivant les types d'applications. Dans le scénario A où le gain en termes de libération des ressources est important grâce à une plus forte proportion d'applications multimédias que de voix, on observe que le taux de blocage est d'autant plus important que l'application est moins prioritaire dans la hiérarchie qui a été établie. En effet, les applications de type *background* subissent le taux de blocage le plus important car lorsqu'une telle session se présente en entrée du réseau, s'il y a congestion, il n'existe pas d'applications qui soient



(a) - scénario A: taux de blocage global

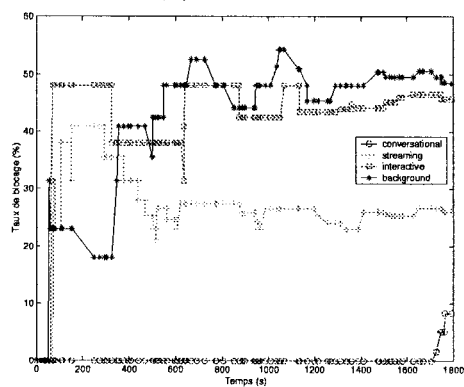


CAC1

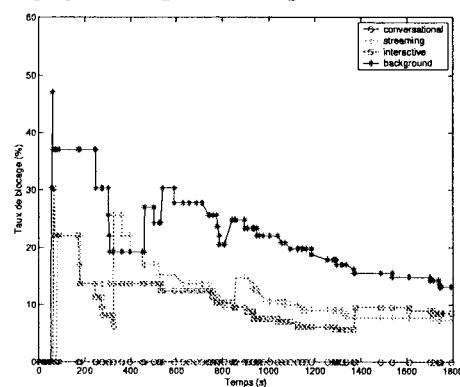


MBCAC

(b) - scénario A: taux de blocage par catégorie d'utilisateur



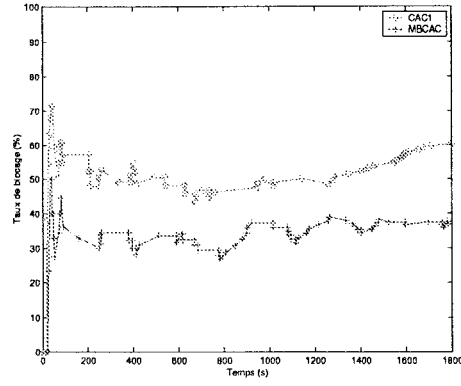
CAC1



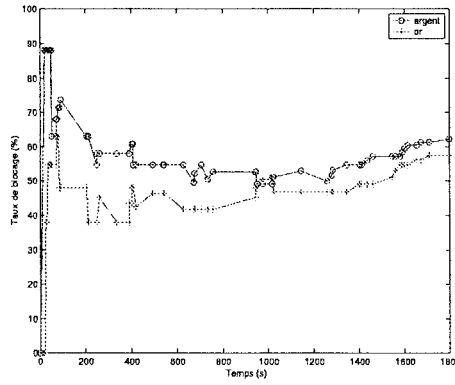
MBCAC

(c) - scénario A: taux de blocage par type d'application

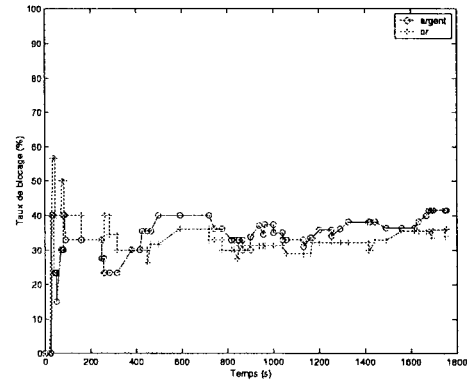
FIG. 4.13 – *NdS* globaux, suivant le profil d'utilisateur et le type d'application, pour le scénario A sous les procédures CAC1 et MBCAC



(a) - scénario B: taux de blocage global

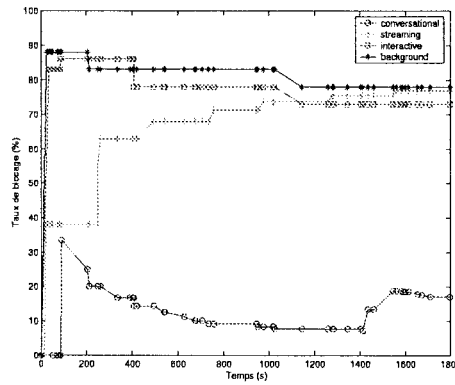


CAC1

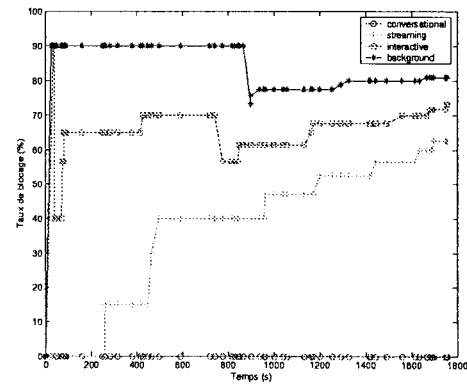


MBCAC

(b) - scénario B: taux de blocage par catégorie d'utilisateur



CAC1



MBCAC

(c) - scénario B: taux de blocage par type d'application

FIG. 4.14 – NdS globaux, suivant le profil d'utilisateur et le type d'application, pour le scénario B sous les procédures CAC1 et MBCAC

hiérarchiquement en dessous et grâce auxquelles on pourrait éventuellement libérer des ressources. Ainsi les applications de voix qui sont très prioritaires sont quasiment toujours acceptées dans le réseau, ce très faible taux de blocage est aussi dû au fait que les sessions voix ont une demande en bande passante relativement faible.

4.5.2 Taux d'occupation à la station de base

La figure 4.15 indique les taux d'utilisation moyens de la capacité dans le sens montant et dans le sens descendant sous chaque procédure, pour le scénario A et le scénario B. L'observation des diagrammes montre que pour le scénario A la voie descendante est beaucoup plus utilisée que la voie montante. Cela est un élément principal des futurs réseaux 3G. Plus le trafic est multimédia et donc asymétrique, plus le lien descendant devient le plus congestionné. Le scénario B appuie également cette conclusion, simplement la différence entre le sens montant et le sens descendant est moins importante car le chargement en trafic voix symétrique est plus important dans ce scénario.

D'autre part, le taux d'occupation moyen est amélioré avec la procédure MBCAC par rapport à CAC1. Le taux d'utilisation moyen dans le sens descendant passe de 77.9% (CAC1) à 80.5% (MBCAC) pour le scénario A, et passe de 56.0% (CAC1) à 59.3% (MBCAC) pour le scénario B. Nous notons cependant que l'écart n'est pas très important. En effet, les deux procédures s'adaptent au chargement du réseau, le MBCAC permet d'admettre plus de sessions en pouvant éventuellement libérer plus de ressources que le CAC1 mais les deux procédures atteignent un taux d'utilisation satisfaisant. Les ressources radios à la station de base sont quand même mieux utilisées avec la procédure MBCAC.

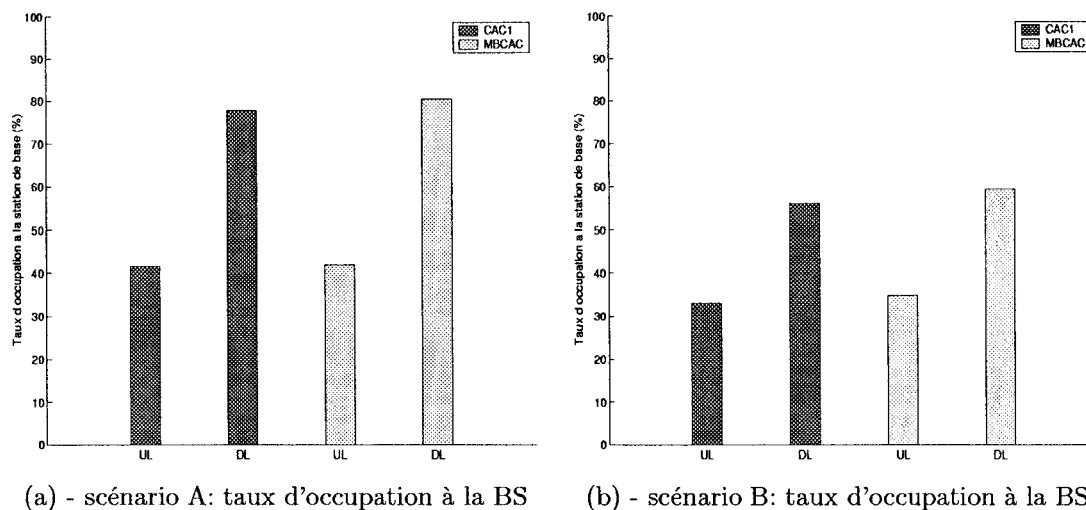


FIG. 4.15 – Taux d'utilisation de la capacité disponible à la BS pour les scénarios A et B sous les procédures CAC1 et MBCAC

TAB. 4.13 – Taux d'utilisation comparés pour les scénarios A et B sous les procédures CAC1 et MBCAC

	scénario A					
	utilisation UL			utilisation DL		
	moy (%)	min (%)	max (%)	moy (%)	min (%)	max (%)
CAC1	41.5	6.8	73.5	77.9	23.6	93.1
MBCAC	41.9	8.8	76.1	80.5	29.1	97.9

	scénario B					
	utilisation UL			utilisation DL		
	moy (%)	min (%)	max (%)	moy (%)	min (%)	max (%)
CAC1	33.0	5.5	65.2	56.0	15.2	91.0
MBCAC	34.7	6.5	67.4	59.3	20.7	99.3

4.5.3 Garantie de QoS

Nous vérifions ici que la garantie sur le délai est bien respectée. Comme attendu, la figure 4.16 montre que le délai est parfaitement respecté pour chacune des procédures, sous chaque scénario. En effet, le bon dimensionnement des *token buckets* et des tailles des files d'attente dans la partie réseau cœur selon les critères de Parekh et Gallager^[16] permettent d'assurer strictement le respect des contraintes de délai.

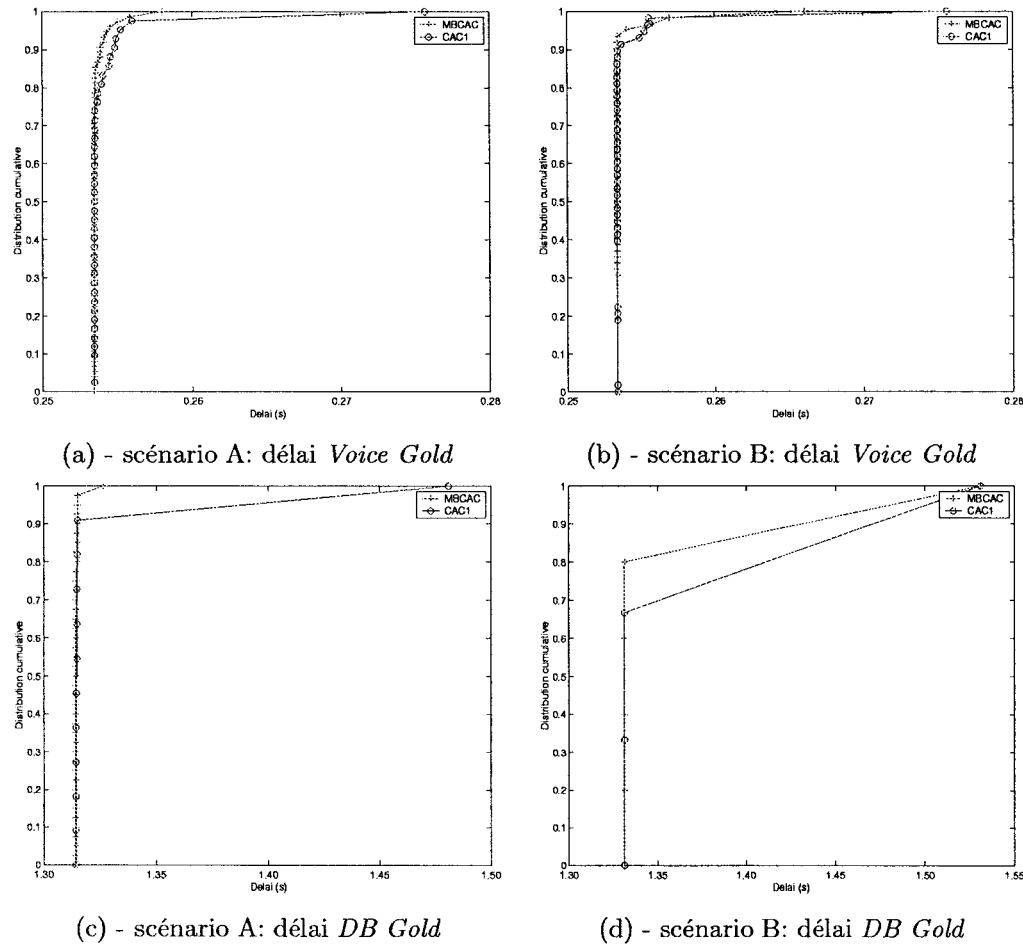


FIG. 4.16 – Distributions cumulatives de la répartition du délai bout en bout pour les applications *Voice Gold* et *DB Gold* sous les procédures CAC1 et MBCAC

4.6 Aide au dimensionnement

La définition d'un algorithme d'admission fait partie intégrante de la problématique de dimensionnement des réseaux de communications. La structure de contrôle et de décision proposée permet d'offrir des garanties de QdS aux sessions en cours tout en effectuant une gestion contrôlée et efficace des ressources. Le dimensionnement du réseau d'accès radio peut donc être évoqué comme suite logique aux résultats de la section précédente. En effet, dans cette section, nous déterminons le taux de blocage global en fonction de différentes valeurs prises pour le rapport $\frac{R}{d_0}$ dans le cas du scénario B sous la procédure MBCAC. Pour cela, nous avons fixé d_0 à 1500 m et nous avons fait prendre à R les valeurs suivantes: 970m, 990m, 995m, 1000m, 1005m, 1010m et 1015m.

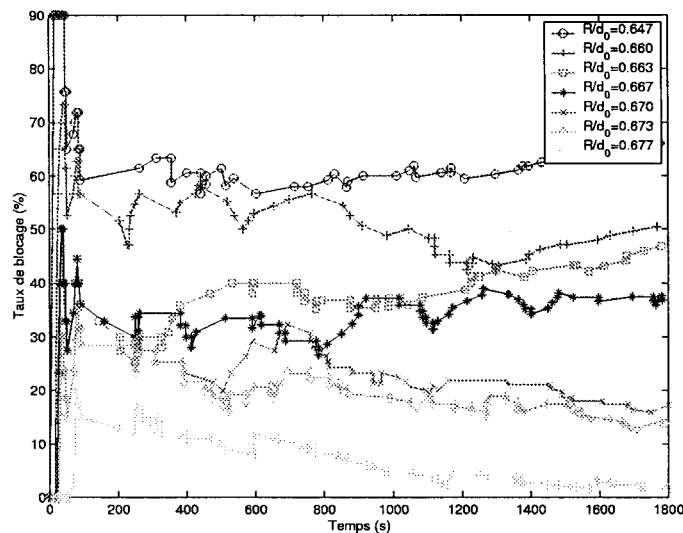


FIG. 4.17 – Taux de blocage en fonction du rayon de la cellule pour le scénario B sous la procédure MBCAC

Il apparaît sur la figure 4.17 que lorsque le rayon de répartition des usagers augmente, le taux de blocage augmente également. En effet, plus les usagers sont supposés être répartis sur une grande surface autour de la station de base, plus les usagers sont éloignés en moyenne, et donc moins on peut en admettre.

Il semble donc intéressant d'utiliser ces résultats pour s'en servir d'aide au dimensionnement radio. Etant donné un plan de stations de base réparties de façon équidistantes

les unes des autres, étant donnée une certaine charge de trafic supposée identique pour chaque cellule, et étant donné un niveau de service NdS à atteindre, c'est-à-dire un taux de blocage à ne pas dépasser, nous pouvons déterminer le rayon de couverture de chaque station de base. Il s'agit de construire un abaque du niveau de service en fonction du rayon R et ce pour une charge de trafic fixe. L'établissement de ces "abaques de service" permet de définir le rayon de couverture maximal pour l'obtention d'un niveau de service donné pour le type de charge de trafic considéré.

4.7 Conclusion

L'environnement de simulation décrit à la section 4.1 a permis d'établir les caractéristiques spécifiques de l'algorithme de contrôle d'admission développé. Dans le cadre des réseaux de troisième génération, nous nous sommes concentrés sur les performances en termes d'utilisation des ressources radios et en termes de taux de blocage, sachant que la partie congestionnante est le lien descendant de l'interface radio.

La procédure MBCAC a ainsi démontré une amélioration des niveaux de service et des taux moyens d'utilisation des ressources radios par rapport à la procédure CAC1. Cette amélioration des performances est rendue possible par un mécanisme de mesure préalablement paramétré en fonction du type de trafic émis. De plus, les performances en termes de qualité de service sont identiques pour chaque procédure, la contrainte de délai est strictement respectée par la présence des *token buckets* pour chaque flot, qui lissent le trafic émis.

L'architecture de contrôle et de décision qui a été définie permet une différenciation des services, une garantie de QoS, une utilisation efficace des ressources et s'intègre dans la problématique du dimensionnement des réseaux de communications mobiles. En effet, le prototype développé à partir des équations CDMA permet l'évaluation de niveaux de service en fonction de caractéristiques spécifiques des cellules radio pour une gamme variée de répartitions de trafic. Le prototype de simulation, tel que développé dans notre mémoire, permet d'obtenir des résultats en vue du déploiement et du dimensionnement du réseau radio, et cela grâce aux équations développées dans le chapitre 2.

CONCLUSION

Dans ce mémoire, notre étude a porté sur la conception et la mise en œuvre d'une architecture de contrôle et de décision susceptible d'être déployée dans les prochaines infrastructures des systèmes cellulaires de troisième génération. Plusieurs décisions ont été prises afin de modéliser le réseau dans son ensemble. Le défi a été de considérer une architecture globale prenant en compte des décisions à la fois sur le réseau radio et sur le réseau cœur, et s'appuyant sur les différents mécanismes qui permettent de fournir une qualité de service adaptée à chaque type d'application. A ce titre, nous avons choisi de traiter chaque session individuellement dans le cadre du modèle à intégration de service et d'intégrer la différenciation des services en introduisant une hiérarchie parmi les différents types d'applications et de profils d'utilisateurs. Nous avons ainsi choisi de mettre en œuvre la politique d'ordonnancement des paquets WFQ pour le service des sessions suivant le mode de commutation par paquets en tout nœud du réseau cœur afin de fournir la qualité de service avec un partage des ressources adéquat. De plus, nous avons développé des formules d'estimation de la capacité radio adaptées au trafic multiservices, afin de fournir une qualité de service satisfaisante pour chaque session sur la partie radio.

L'objectif du mémoire est d'élaborer les performances d'un contrôle d'admission à base de mesures s'appuyant sur l'architecture originale préalablement établie, et de les comparer avec celles obtenues par un contrôle d'admission se basant sur des descripteurs de trafic à priori. L'étude met l'accent sur les performances radio car il s'agit de la partie congestionnante dans les futurs réseaux de troisième génération, tout en s'assurant de la garantie du service sur le réseau bout en bout. Suite à ces expérimentations, nous pouvons valider les formules de capacité radio développées et s'en servir pour fournir un outil d'aide au dimensionnement du réseau d'accès radio.

La démarche suivie fut la suivante:

- S'appuyant sur la mise en œuvre de la discipline WFQ et d'une estimation des ressources nécessaires à la satisfaction des contraintes de QoS, nous proposons un

algorithme de contrôle d'admission adaptatif à base de mesures sur le réseau cœur afin d'obtenir une garantie de la qualité de service et un taux efficace d'utilisation des ressources.

- Le réseau d'accès radio, partie clé des futurs réseaux radiomobiles, a fait l'objet d'une étude approfondie. Dans ce mémoire nous finalisons des équations d'estimation de la capacité radio CDMA dans un contexte multiservices, et prenant en compte le mécanisme de *soft handover*. Ces équations forment le noyau d'un algorithme de contrôle d'admission sur le réseau radio permettant la satisfaction des contraintes de QoS de toute session entrant sur le réseau d'accès radio.
- Intégrant les concepts de la partie réseau cœur et de la partie radio, une architecture de contrôle et de décision originale bout en bout est développée. Elle s'appuie sur des mesures du débit émis par chaque source de trafic et a pour objectif de réaliser une gestion efficace des ressources radios et satisfaire les contraintes de QoS bout en bout du plus grand nombre de sessions.
- Implantant cette architecture sous l'environnement de modélisation de réseaux OPNET, le prototype développé permet d'évaluer les performances et caractéristiques de l'algorithme de contrôle d'admission, et d'évaluer l'influence des décisions prises en termes de conception et de déploiement du réseau d'accès radio.

Les résultats obtenus nous permettent plusieurs conclusions:

- Dans le contexte des réseaux 3G, l'algorithme d'admission à base de mesures présente des performances variables en fonction des paramètres de mesure utilisés. Etant donné la nature du trafic émis dans notre simulateur, il est possible après plusieurs simulations tests, de donner des valeurs optimales aux paramètres de mesure, faisant en sorte que l'on obtient le taux de blocage le plus faible possible, tout en assurant que le débit mesuré reste un bon indicateur du comportement futur de la source. Ainsi le phénomène de blocage et le NdS qui en résulte constituent l'élément caractéristique de la procédure de contrôle d'admission.
- Du point de vue de ses caractéristiques, l'algorithme de contrôle d'admission présen-

te les propriétés attendues lors de sa conception. Il est adaptatif, sélectif et s'ajuste à l'évolution de la charge du réseau en privilégiant lorsque nécessaire les sessions de hautes priorités par rapport aux sessions de priorités faibles lors de l'admission. De plus, il garantit une borne supérieure sur le délai subi par chaque application et satisfait la demande en bande passante de chaque session.

- En termes de performances, nous avons mis en évidence les avantages d'un algorithme à base de mesures dans le contexte des réseaux 3G. Par rapport à une procédure basée sur des descripteurs de trafic à priori, l'algorithme à base de mesures montre un gain notable en termes de taux de blocage, et une meilleure utilisation des ressources radios, notamment sur le lien descendant qui est le lien critique avec du trafic multimédia asymétrique.
- En ce qui concerne les formules de capacité radio développées, une analyse détaillée a mis en évidence la possibilité d'utiliser le prototype de simulation pour une aide au dimensionnement du réseau d'accès radio. Pour des répartitions variées de trafic, les expérimentations permettent de déterminer les rayons de couverture maximaux des différentes stations de base afin d'atteindre un certain niveau de service. Cela ouvre la voie à de multiples approches pour le dimensionnement des futurs réseaux de communications mobiles.

Toute modélisation disposant de limites et de restrictions, la modélisation présentée dans le cadre de notre étude présente les siennes :

- Dans notre étude, le processus de mesure permet uniquement d'effectuer des réévaluations à la baisse des montants de ressources allouées. Le contrôle d'admission à base de mesures évalue en ligne les ressources occupées, et les sessions pourraient donc faire l'objet à la fois de fluctuations à la hausse et à la baisse en fonction des mesures faites tout au long de la période d'activité. Ceci permettrait d'obtenir un gain de performances encore meilleur en termes de nombre de sessions concurrentes admises et en termes d'utilisation des ressources, au pris toutefois de violations occasionnelles de garanties de service.

- Le point soulevé ci-dessus est fortement lié à la nature du trafic que l'on considère. En effet, pour envisager des fluctuations à la hausse et à la baisse, il s'agit de modéliser du trafic fortement variable pour y voir un gain de performances, et également de mener une étude poussée sur le mécanisme de mesure qui doit être le plus précis possible. Dans notre étude, le trafic modélisé est moyennement variable et le mécanisme de mesure est le modèle simple proposé par Jamin *et al.*^[35]. Il serait intéressant de mener un approfondissement sur la manière de rendre ces mesures encore plus fiables et de choisir les paramètres de mesure optimaux.
- Les formules de capacité radio comportent encore certaines limitations. D'abord nous supposons que le trafic est le même dans chaque cellule radio ce qui n'est pas le cas dans la réalité. D'autre part, pour le lien montant, l'hypothèse que les interférences inter-cellulaires sont une fraction de la puissance totale reçue à la station de base est trop restrictive. Cela engendre l'impossibilité de faire du dimensionnement radio avec les équations dans le sens montant. Il s'agit donc de relier la quantité d'interférences inter-cellulaires en fonction du pourcentage d'utilisateurs en *soft handover*, ce dernier étant lui même relié au rayon sur lequel sont répartis les utilisateurs.

RÉFÉRENCES

- [1] J. Sanchez and M. Thioune, ed., *UMTS: services, architecture et WCDMA*. Paris: Editions Hermès, 2001. Collection Réseaux et télécommunications.
- [2] C. Perkins, *cdma2000 Wireless IP Network Standard*. Network Working Group, October 1996. Internet RFC 2002.
- [3] TIA, *IP Mobility Support*. Telecommunications Industry Association, September 2002. TIA/IS-835-B committee TR-45.6.
- [4] H. Holma and A. Toskala, ed., *WCDMA for UMTS - Radio Access for Third Generation Mobile Communications*. Chichester, England: Wiley, 2001. Revised Edition.
- [5] TIA, *Physical Layer Standard for cdma2000 Spread Spectrum Systems*. Telecommunications Industry Association, May 2002. TIA/EIA/IS-2000.2-C.
- [6] 3GPP, “3rd Generation Partnership Project. Technical Specification Group Services and Systems Aspects; QoS Concept and Architecture (Release 1999),” Tech. Rep. TS 23.107 v3.2.0, <http://www.3gpp.org>, 2000.
- [7] O. Troeung, “Un algorithme de contrôle d’admission pour la garantie de qualité de service dans le réseau de cœur des systèmes de troisième génération,” Master’s thesis, Ecole Polytechnique de Montréal, December 2001.
- [8] B. Jaumard, R. Malhamé and O. Troeung, “A Dimensioning Simulator for Multimedia 3G Networks - Core Network Component,” tech. rep., Ecole Polytechnique de Montréal, February 2001. Ericsson Contract #1, Deliverable #2.
- [9] IETF. <http://www.ietf.org>.
- [10] S. B. R. Braden, L. Zhang, S. Herzog, and S. Jamin, *Resource ReServation Protocol (RSVP) – Version 1 Functional Specification*. Network Working Group, September 1997. Internet RFC 2205.
- [11] J. Kurose, “Open issues and challenges in providing QoS guarantees in high speed networks,” *CCR*, vol. 23(1), pp. 6–15, January 1993.

- [12] C. Aras, J. Kurose, D. Reeves and H. Schulzrinne, "Real-time communication in packet-switched networks"," *Proceedings of the IEEE*, vol. 82(1), pp. 122–139, January 1994.
- [13] J. Nagle, "On packet switches with infinite storage," *IEEE Transactions on Communication*, vol. 35(4), pp. 435–438, April 1987.
- [14] A. Demers, S. Keshav and S. Shenker, "Analysis and Simulation of a Fair Queuing Algorithm," *Proceedings ACM SIGCOMM'89*, pp. 3–12, September 1989.
- [15] A. Parekh and R. 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, June 1993.
- [16] A. Parekh and R. 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, April 1994.
- [17] S. J. Golestani, "A self-clocked fair queuing scheme for broadband applications," *Proceedings of INFOCOMM'94*, pp. 636–646, June 1994.
- [18] J.C.R. Bennet and H. Zhang, "WF²Q: Worst-case fair weighted fair queueing," *Proceedings of INFOCOMM'96*, pp. 1d.3.1–1d.3.9, 1996.
- [19] P. Goyal, H.M. Vin and H.Cheng, "Start-time Fair Queueing: A scheduling algorithm for integrated services packet switching networks," *Proceedings of SIGCOMM'96*, pp. 157–168, 1996.
- [20] L. Zhang, "Virtual Clock: A new traffic control algorithm for packet switching networks," *Proceedings of SIGCOMM'90*, pp. 19–29, September 1990.
- [21] D. Levine, M. Naghshineh and I. Akildyz, "A Ressource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept," *IEEE/ACM Transactions on Networking*, vol. 5(1), pp. 1–12, February 1997.
- [22] K. Lee, "Supporting Mobile Multimedia in Integrated Services Networks," *Wireless Networks, ACP Press*, vol. 2, August 1996.

- [23] A.M. Viterbi and A.J. Viterbi, "Erlang capacity of a power controlled CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 11(6), pp. 892–900, August 1993.
- [24] Z. Liu and M. El Zarki, "SIR-based Call Admission Control for DS-CDMA Cellular Systems," *Wireless Networks, ACP Press*, vol. 12(4), pp. 638–644, May 1994.
- [25] M.A. Landolsi, V.V. Veeravalli and N. Jain, "New results on the reverse link capacity of CDMA cellular networks," *Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th*, vol. 3, pp. 1462–1466, 1996.
- [26] K. Sipila, K.-C. Honkasalo, J. Laiho-Steffens and A. Wacker, "Estimation of capacity and required transmission power of WCDMA downlink based on a downlink pole equation," *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 2, pp. 1002–1005, August 2000.
- [27] K. Hiltunen and R. de Bernardi, "WCDMA downlink capacity estimation," *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 2, pp. 992–996, 2000.
- [28] C. Cordier and S. Ortega, "On WCDMA downlink multiservice coverage and capacity," *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, vol. 4, pp. 2754–2758, 2001.
- [29] S. Akhtar, S.A. Malik and D. Zeghlache, "Prioritized admission control for mixed services in UMTS WCDMA networks," *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 1, pp. B-133–B-137, 2001.
- [30] W.S. Jeon and D.G. Jeong, "Admission control of multimedia calls in CDMA mobile systems," *Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd*, vol. 2, pp. 886–890, 2001.
- [31] C. P. S. Shenker and R. Guerin, *Specification of Guaranteed Quality of Service*. Network Working Group, September 1997. Internet RFC 2212.
- [32] H. Zhang, "Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks," *Proceedings of IEEE*, vol. 83(10), pp. 1374–1396, October

1995.

- [33] Z.L. Zhang, Z. Liu, J. Kurose and D. Towsley, "Call Admission Control Schemes under the Generalized Processor Sharing Scheduling," *Rapport de recherche de l'INRIA*, 1995.
- [34] J. Wroclawski, *Specification of the Control-Load Network Element Service. RFC2211*. Network Working Group, September 1997. Internet RFC 2211.
- [35] S. Jamin, P.B. Danzig, S.J. Shenker and L. Zhang, "A Measurement-Based Admission Control Algorithm for Integrated Service Packet Networks," *IEEE/ACM Transactions on Networking*, vol. 5(1), pp. 56–70, February 1997.
- [36] S. Jamin, S.J. Shenker and P.B. Danzig, "Comparison of Measurement-based Admission Control Algorithms for Controlled-Load Service," *Proceedings of IEEE INFOCOM'97, Kobe, Japan*, April 1997.
- [37] S. Jamin and S. Shenker, "Measurement-based Admission Control Algorithms for Controlled-load Service: A Structural Examination," *University of Michigan, Technical Report CSE-TR-333-97*, April 1997.
- [38] L. Breslau, S. Jamin and S. Shenker, "Comments on the Performance of Measurement-based Admission Control Algorithms," *Proceedings of IEEE INFOCOM 2000*, 2000.
- [39] C. Casetti, J. Kurose and D. Towsley, "A new Algorithm for Measurement-based Admission Control in Integrated Services Packet Networks," *IFIP Protocols for High Speed Networks '96*, 1996.
- [40] D. Liu and Y. Cai, "A Heuristic Approach for Measurement-Based Admission Control with Variable-Size Window," *Proceedings of the IEEE Global Communications Conference*, pp. 3628–3632, November 2001.
- [41] J. Qiu and E. Knightly, "Measurement-based Admission Control with Aggregate Traffic Enveloppes," *IEEE/ACM Transactions on Networking*, vol. 9(2), pp. 199–210, April 2001.
- [42] S. Floyd, "Comments on Measurement-based Admissions Control for Controlled-load Services," *Technical Report, Lawrence Berkeley National Laboratory*, p. 20, July

1996.

- [43] M. Grossglauser and D.N.C. Tse, "A Framework for Robust Measurement-Based Admission Control," *IEEE/ACM Transactions on Networking*, vol. 7(3), pp. 293–309, 1999.
- [44] S. Tsang-Kwong-U, "Evaluation du problème de déficit de puissance sur le lien descendant dans les réseaux cellulaires CDMA," Master's thesis, Ecole Polytechnique de Montréal, Department of Computer and Electrical Engineering, December 2000.
- [45] B. Jaumard, F. Lupien, R. Malhamé and S. Esposito, "Call Admission Control Policies and Dimensioning of Multimedia 3G Networks - Simulation Plan," tech. rep., Ecole Polytechnique de Montréal, April 2002. Ericsson Contract #1, Deliverable #2, Revised.
- [46] Opnet Technologies Inc. <http://www.opnet.com>.
- [47] Opnet Technologies Inc, "Standard Network Applications," tech. rep., Opnet, July 2001.
- [48] H. ElGebaly, "Characterization of Multimedia Streams of an H.323 Terminal," Tech. Rep. Q2 1998, Intel Architecture Labs, http://www.intel.com/technology/itj/q21998/articles/art_5.htm, 1998.
- [49] B. Lind, "Application Traffic Model for UMTS Services," tech. rep., Ericsson, January 2000.
- [50] S. Madan and M.M. Diez, "IS-2000 Forward Link Simulation-Performance Results," tech. rep., Ericsson, October 2000. Ericsson Documentation.
- [51] B.J. Kang, "Performance Evaluation of Reverse Link for Speech and Data Traffic in CDMA-Based IMT-2000," *IEEE International Conference on Communications*, vol. 3, pp. 1409–1413, 1999.
- [52] B.J. Kang and S.Y. Kim, "Capacity Calculation of Reverse Link in CDMA-Based IMT2000," *Proceedings of the IEEE Region 10 Conference*, vol. 2, pp. 1251–1254, 1999.
- [53] W.C.Y. Lee, ed., *Mobile Cellular Telecommunications*. McGraw-Hill, Inc., 1995. Second Edition.

Annexe I

$\frac{E_b}{N_t}$ requis en fonction du RAB alloué

Des mesures ont été faites et reportées dans la littérature^{[50][51][52]} permettant de connaître la valeur de $\frac{E_b}{N_t}$ requise à la station de base pour le sens montant, et au mobile pour le sens descendant, en fonction de la valeur du FER et du débit R alloués à la session.

Nous donnons ici les valeurs utilisées dans le cadre de notre étude. Ces valeurs sont celles obtenues pour un usager circulant à 3km/h (piéton). Les mesures sont faites pour une configuration des liens radios fonctionnant sur les débits discrets suivants: 9600bps, 19200bps, 38400bps, 76800bps, 153600bps, 307200bps.

I.1 Abaques pour le sens descendant

TAB. I.1 – γ_s^{DL} (en dB) en fonction de FER_s^{DL} et R_s^{DL}

		FER	
		1%	5%
R (bps)	9600	3.7	2.8
	19200	3.7	2.88
	38400	3.58	2.95
	76800	3.78	3.12
	153600	3.91	3.28
	307200	3.98	3.55

I.2 Abaques pour le sens montant

Dans le sens montant, le rapport requis énergie par bit sur densité spectrale du bruit est de 2dB en dessous de celui requis pour le sens descendant.

TAB. I.2 – γ_s^{UL} (en dB) en fonction de FER_s^{UL} et R_s^{UL}

		FER	
		1%	5%
R (bps)	9600	1.7	0.8
	19200	1.7	0.88
	38400	1.58	0.95
	76800	1.78	1.12
	153600	1.91	1.28
	307200	1.98	1.55

Annexe II

Calcul des coefficients d'atténuation pour les formules de capacité radio dans le sens descendant

II.1 Modèle de propagation

Nous considérons dans cette étude un modèle simple de propagation, comme décrit par Lee^[53]. L'onde radio est transmise par une antenne de hauteur h_a vers un usager mobile de hauteur h_m . Soit P_0 la puissance de l'onde émise par l'antenne. Notons P_r la puissance reçue au mobile. Si l'on considère que le mobile reçoit seulement l'onde directe et l'onde réfléchie, alors on a:

$$P_r = P_0 \left(\frac{1}{4\pi d/\lambda} \right)^2 |1 + a_\nu e^{j\Delta\Phi}|^2 \quad (\text{II.1})$$

avec

- $a_\nu = -1$ le coefficient de réflexion
- d la distance entre l'antenne et le mobile
- λ la longueur d'onde
- $\Delta\Phi \approx \frac{2\pi}{\lambda} \frac{2h_a h_m}{d}$ le déphasage entre l'onde directe et l'onde réfléchie

ce qui donne, en considérant le déphasage Φ petit, le coefficient d'atténuation suivant:

$$\frac{P_0}{P_r} = \frac{d^4}{h_a^2 h_m^2} \quad (\text{II.2})$$

II.2 Calcul des coefficients $E[a_{m_s,0}]$ et $E[\frac{a_{m_s,0}}{a_{m_s,i}}]$

Les calculs présentés ici s'appuient sur le travail de Steeve Tsang-Kwong-Hu^[44]. Soit d la variable qui représente la distance entre le mobile et sa station de base. La position du mobile dans sa cellule est repérée par ses coordonnées polaires (d, θ) , la station de base prenant en charge le mobile étant l'origine du repère. Les stations de base sont réparties

de façon équidistantes et on note d_0 la distance séparant deux stations de base voisines. La figure II.1 illustre cette configuration.

• **Distribution des usagers uniforme suivant la distance à la station de base**

On suppose que les usagers sont répartis uniformément suivant la distance à la station de base, et ce sur une surface circulaire de rayon R (on considère donc une cellule circulaire pour ces calculs). La densité de probabilité de la position du mobile est donc:

$$f_d(d, \theta) = \frac{1}{2\pi R}$$

$a_{m_s,0} = \frac{d^4}{K}$ avec $K = h_a^2 h_m^2$ d'après la section II.1 (On prendra dans les calculs une hauteur d'antenne de 30m et une hauteur pour le mobile de 2m). Ainsi on peut calculer:

$$E[a_{m_s,0}] = \int_0^R \int_0^{2\pi} \frac{1}{K} d^4 f_d(d, \theta) dd d\theta = \frac{1}{K} \frac{1}{5} R^4. \quad (\text{II.3})$$

Notons θ_i l'angle entre la station de base du mobile et la station de base voisine i , $i \in \{1, \dots, 6\}$. On a alors $a_{m_s,i} = \frac{1}{K} (d^2 + d_0^2 - 2dd_0 \cos(\theta - \theta_i))^2$. Ainsi on peut calculer

$$E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] = \int_0^{2\pi} \int_0^R \frac{d^4}{(d^2 + d_0^2 - 2dd_0 \cos(\theta - \theta_i))^2} f_d(d, \theta) d\theta dd \quad (\text{II.4})$$

$\theta_i \in \{\frac{\pi}{6}, \frac{\pi}{2}, \frac{5\pi}{6}, \frac{7\pi}{6}, \frac{3\pi}{2}, \frac{11\pi}{6}\}$ et on rappelle que l'équation II.4 donne le même résultat pour chacune de ces valeurs de θ_i car la répartition des usagers est uniforme sur la cellule.

Ainsi $E[\frac{a_{m_s,0}}{a_{m_s,i}}]$ dépend uniquement de la valeur de d_0 et de R . Pour résoudre l'intégrale le logiciel Maple a été employé et je remercie Christophe Meyer qui a établi le résultat suivant :

$$E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] = -1 + \frac{7}{4\left(\frac{R^2}{d_0^2} - 1\right)} + \frac{1}{2\left(\frac{R^2}{d_0^2} - 1\right)^2} + \frac{9d_0}{8R} \ln\left(\frac{1 + \frac{R}{d_0}}{1 - \frac{R}{d_0}}\right) \quad (\text{II.5})$$

• **Distribution des usagers uniforme dans l'espace de la cellule**

On suppose que les usagers sont répartis uniformément dans l'espace de la cellule, soit

sur une surface circulaire de rayon R (on considère donc à nouveau une cellule circulaire pour ces calculs). La densité de probabilité de la position du mobile est donc:

$$f_d(d, \theta) = \frac{d}{\pi R^2}$$

Ainsi on peut calculer:

$$E[a_{m_s,0}] = \int_0^R \int_0^{2\pi} \frac{1}{K} d^4 f_d(d, \theta) d\theta dd = \frac{1}{K} \frac{1}{3} R^4. \quad (\text{II.6})$$

Puis:

$$E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] = \int_0^{2\pi} \int_0^R \frac{d^4}{(d^2 + d_0^2 - 2dd_0 \cos(\theta - \theta_i))^2} f_d(d, \theta) d\theta dd \quad (\text{II.7})$$

$\theta_i \in \{\frac{\pi}{6}, \frac{\pi}{2}, \frac{5\pi}{6}, \frac{7\pi}{6}, \frac{3\pi}{2}, \frac{11\pi}{6}\}$ et on rappelle que l'équation II.7 donne le même résultat pour chacune de ces valeurs de θ_i car la répartition des usagers est uniforme sur la cellule.

Ainsi $E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right]$ dépend uniquement de la valeur de d_0 et de R . Le logiciel Maple a été utilisé pour résoudre l'intégrale, le calcul a été fait par Christophe Meyer et a donné le résultat suivant :

$$E\left[\frac{a_{m_s,0}}{a_{m_s,i}}\right] = -1 - 4 \frac{d_0^2}{R^2} \ln\left(1 - \frac{R^2}{d_0^2}\right) - \frac{d_0^2}{R^2} \frac{9}{2\left(1 - \frac{R^2}{d_0^2}\right)} + \frac{1 + \frac{d_0^2}{R^2}}{2\left(1 - \frac{R^2}{d_0^2}\right)^2} + 4 \frac{d_0^2}{R^2} \quad (\text{II.8})$$

II.3 Calcul des coefficients $a_{m^*,0}$ et $a_{m^*,i}$

On considère un mobile m^* en *soft handover*, placé au milieu entre sa station de base et la station de base voisine qui le prend aussi en charge. On veut calculer $\sum_{i=1}^6 \frac{a_{m^*,0}}{a_{m^*,i}}$. Supposons que le mobile est desservi par sa station de base et par la station de base indice 1. D'après la figure II.1 on a:

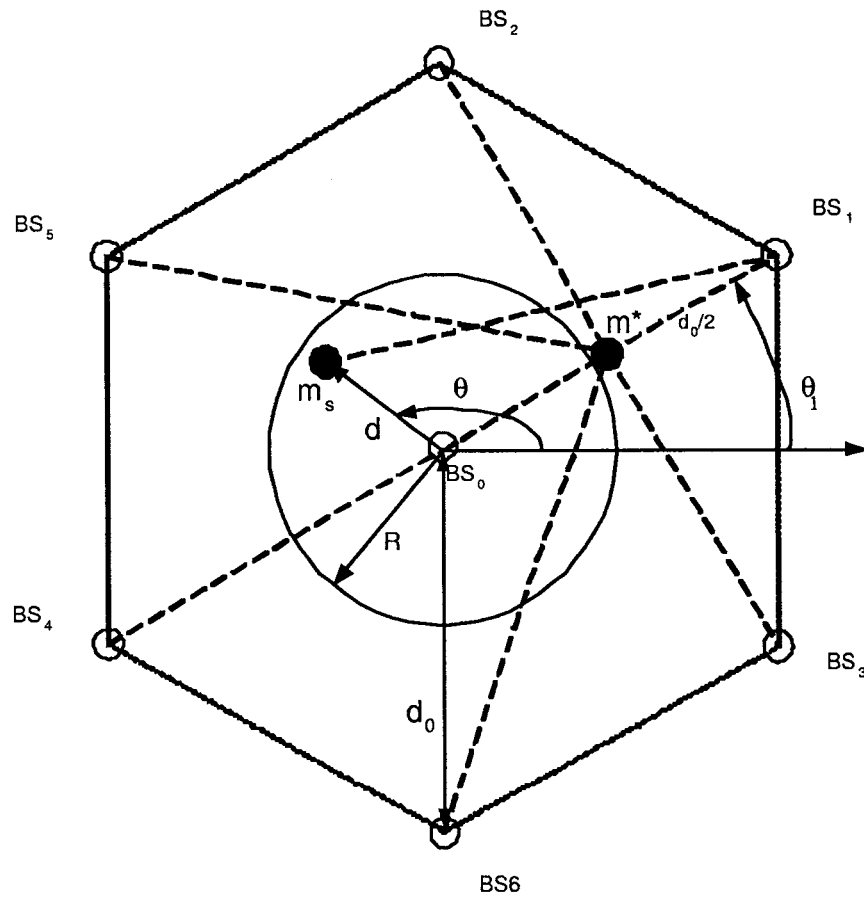


FIG. II.1 – Schéma des cellules pour le calcul des coefficients d'atténuation

$$a_{m*,0} = \frac{1}{K} \left(\frac{d_0}{2} \right)^4$$

$$a_{m*,1} = \frac{1}{K} \left(\frac{d_0}{2} \right)^4$$

$$a_{m*,2} = \frac{1}{K} \left(\frac{3d_0}{2\sqrt{3}} \right)^4$$

$$a_{m*,3} = \frac{1}{K} \left(\frac{3d_0}{2\sqrt{3}} \right)^4$$

$$\begin{aligned}
 a_{m*,4} &= \frac{1}{K} \left(\frac{3d_0}{2} \right)^4 \\
 a_{m*,5} &= \frac{1}{K} \left(d_0^2 + \left(\frac{3d_0}{2\sqrt{3}} \right)^2 \right)^2 \\
 a_{m*,6} &= \frac{1}{K} \left(d_0^2 + \left(\frac{3d_0}{2\sqrt{3}} \right)^2 \right)^2
 \end{aligned}$$

Ainsi pour tout mobile en *soft handover* placé au milieu entre ses deux stations de base on a:

$$\begin{aligned}
 \sum_{i=1}^6 \frac{a_{m*,0}}{a_{m*,i}} &= \left(\frac{d_0}{2} \right)^4 \left(\frac{1}{\left(\frac{d_0}{2} \right)^4} + \frac{2}{\left(\frac{3d_0}{2\sqrt{3}} \right)^4} + \frac{2}{\left(d_0^2 + \left(\frac{3d_0}{2\sqrt{3}} \right)^2 \right)^2} + \frac{1}{\left(\frac{3d_0}{2} \right)^4} \right) \\
 &= 1,2346 \quad \forall d_0
 \end{aligned}$$

Annexe III

Les mécanismes de retransmission de l'ARQ

Les trois mécanismes de l'ARQ sont le *stop and wait*, le *go-back-N* et celui que nous utilisons dans notre modèle : le *selective repeat* (SRP).

III.1 *Stop and wait*

La stratégie la plus simple est appelée *stop and wait* : l'émetteur s'arrête après chaque émission d'une trame et attend soit un message d'acquiescement ACK ou NACK de la part du récepteur, soit l'expiration d'une temporisation. Si un acquiescement positif est reçu (ACK), l'émetteur envoie la trame suivante. Si un acquiescement négatif est reçu (NACK), la dernière trame émise est retransmise. Les performances de ce mécanisme dépendent fortement du temps d'aller-retour, c'est-à-dire du temps passé entre l'envoi d'une trame et la réception de l'ACK/NACK correspondant. Ce protocole est efficace mais le problème principal réside dans le fait que son implémentation nécessite une taille de mémoire tampon importante car les trames peuvent être émises et reçues de façon déséquencées.

III.2 *Go-back-N*

Dans le protocole *go-back-N*, l'émetteur peut envoyer continuellement ses trames tant qu'il en a à transmettre, sans attendre l'ACK. A la réception d'un ACK, l'émetteur peut identifier la dernière trame envoyée et bien reçue par le récepteur et qui n'est pas forcément la dernière trame envoyée (envoi par anticipation). En revanche à la réception d'un NACK la dernière trame envoyée et non acquittée et toutes celles qui suivent sont réémises. Ainsi certaines trames peuvent être envoyées plusieurs fois sans être sujettes à une erreur de transfert. De plus l'émetteur doit garder en mémoire, dans des tampons, les trames envoyées et non encore acquittées positivement. A la réception d'un ACK, l'émetteur peut libérer une place dans ses tampons. Un inconvénient de ce protocole est

que certains messages reçus sans erreurs sont retransmis quand même, augmentant la quantité de données transmises sur les canaux de communications.

III.3 Répétition sélective

Dans la méthode de répétition sélective, schématisée sur la figure III.1, seuls les messages contenant des erreurs à la réception sont retransmis. Dans ce cas, il faut accepter de recevoir les messages de façon déséquencée. Ce protocole, contrairement aux deux précédents, doit utiliser des ACK/NACK avec un mécanisme de numérotation des trames. Les messages ACK/NACK contiennent le numéro de la trame concernée par l'acquiescement. Lors de la réception d'un NACK de la trame n , seule la trame n acquittée négativement est renvoyée. Le récepteur doit faire un réordonnancement des trames reçues pour les transmettre dans l'ordre au récepteur. Enfin, l'émetteur doit garder en mémoire l'ensemble des trames envoyées et non encore acquittées positivement. Cette technique montre des performances meilleures que les deux précédentes mais son implantation est plus complexe, nécessitant la gestion des tampons et une procédure de reséquencement des messages.

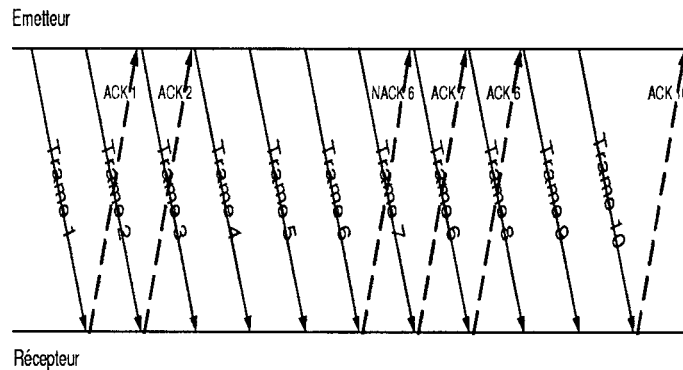


FIG. III.1 – *Protocole de répétition sélective*

Annexe IV

Les algorithmes d'admission CAC0 et CAC1

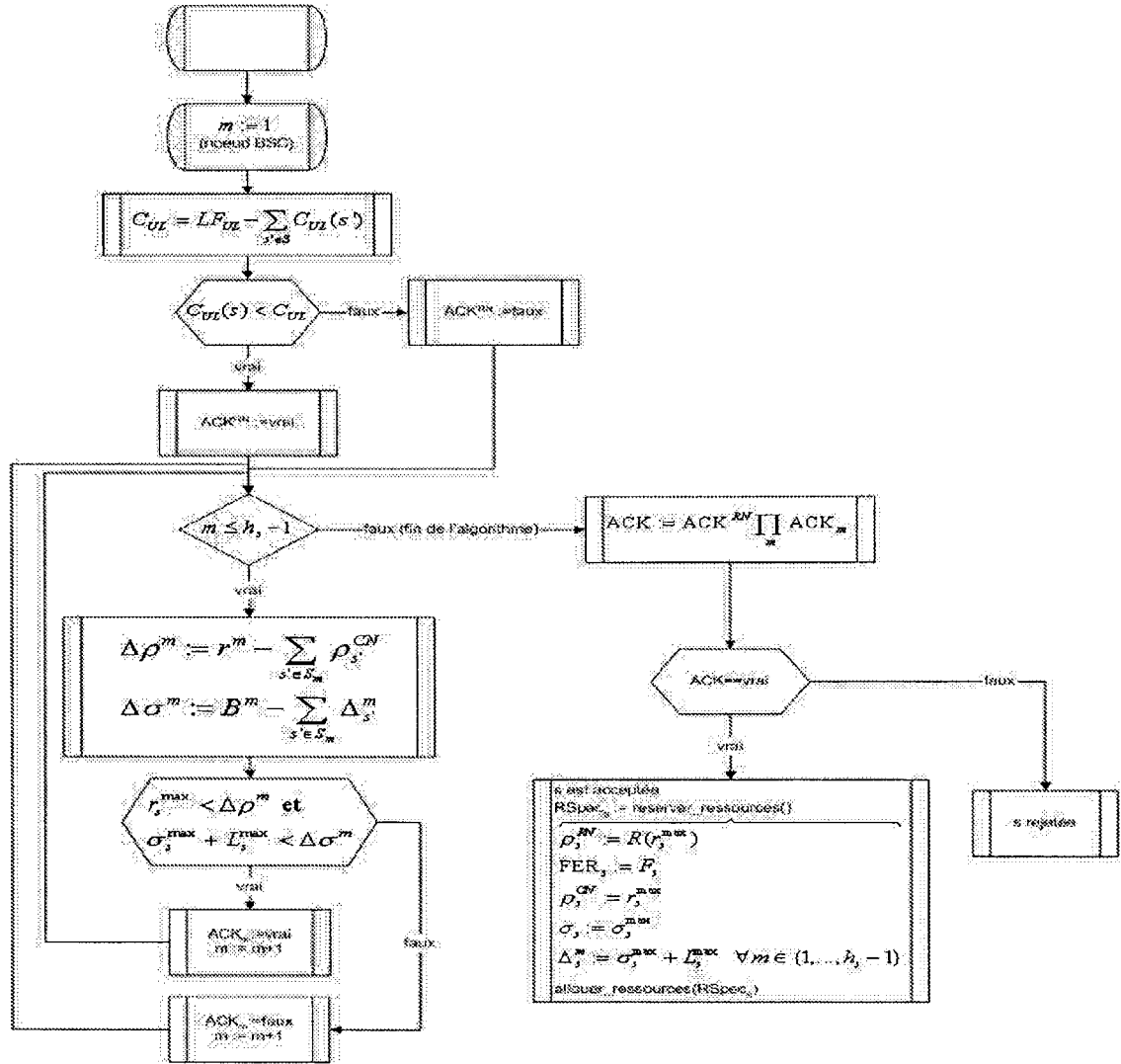


FIG. IV.1 – Algorithme CAC0 sens montant

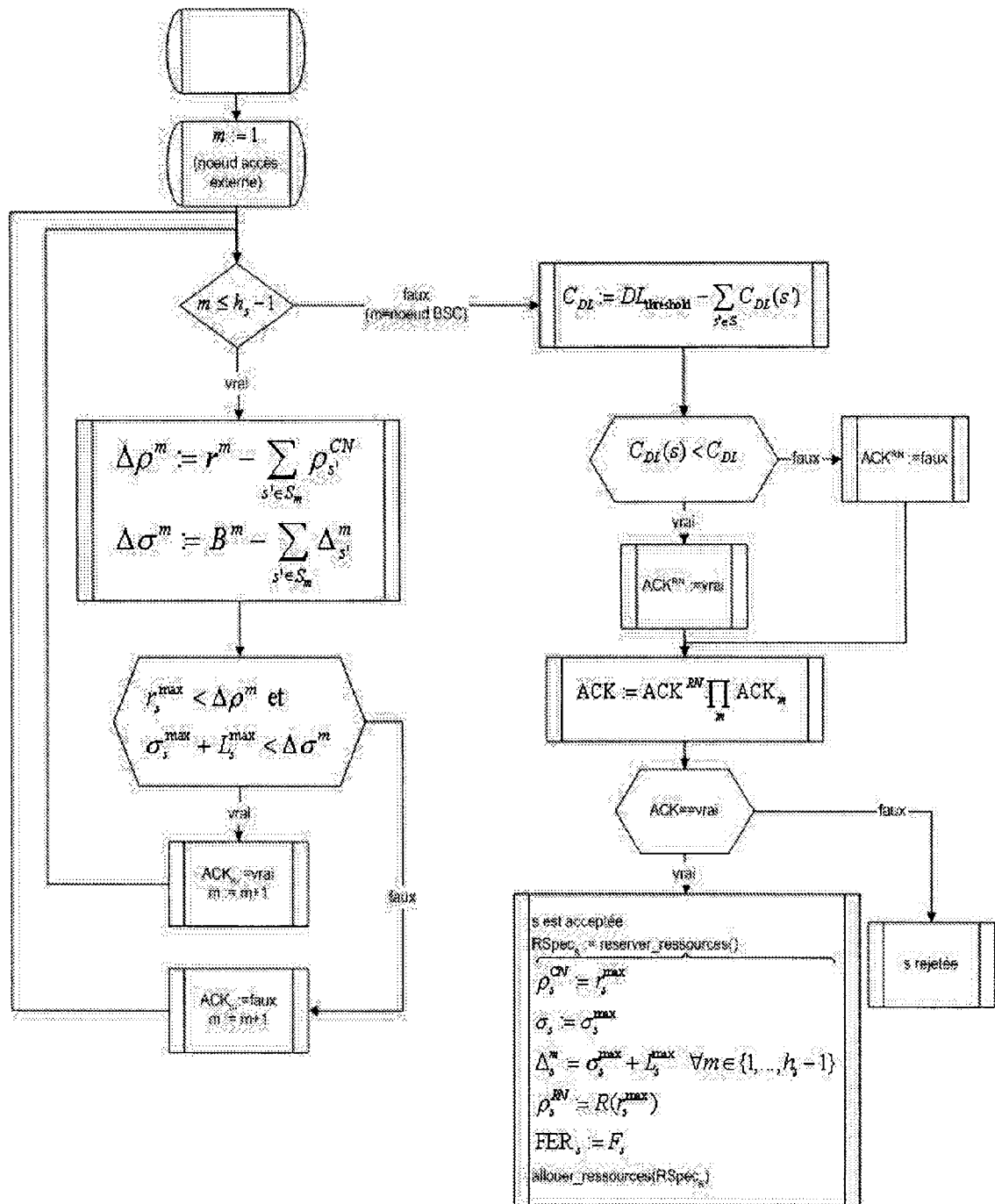


FIG. IV.2 – Algorithme CACO sens descendant

FIG. IV.3 – *Algorithme CAC1 sens montant*

