



HAL
open science

Gestion des ressources des réseaux mobiles de nouvelle génération par rapport à la mobilité des utilisateurs

Khalil Ibrahim

► **To cite this version:**

Khalil Ibrahim. Gestion des ressources des réseaux mobiles de nouvelle génération par rapport à la mobilité des utilisateurs. Autre. Université d'Avignon, 2009. Français. NNT : 2009AVIG0169 . tel-00453644

HAL Id: tel-00453644

<https://theses.hal.science/tel-00453644>

Submitted on 5 Feb 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

présentée pour obtenir le grade de Docteur en Sciences de l'Université d'Avignon et des Pays de Vaucluse France & Université Mohammed V-Agdal Rabat - Maroc

SPÉCIALITÉ : Informatique

École Doctorale 166 « Information Structures Systèmes »
Laboratoire d'Informatique d'Avignon (UPRES No 4128)

Gestion des ressources des réseaux mobiles de nouvelle génération par rapport à la mobilité des utilisateurs

par

Khalil IBRAHIMI

Soutenue publiquement le 20 novembre 2009 devant un jury composé de :

M. BOUYAKHF El-Houssine	Professeur, LIMIARF/FSR, Rabat	Président
M. ALTMAN Eitan	Dir. de recherche, INRIA, Sophia Antipolis	Directeur
M. EL-AZOUZI Rachid	Maître de Conférences, LIA, Avignon	Co-directeur
M. BEN-OTHTMAN Jalel	Maître de Conférences-HDR, Prism, Versailles	Rapporteur
M. ELKOUCH Rachid	Professeur, INPT, Rabat	Rapporteur
M. BENKHALIFA Mohamed	Professeur, Université Mohammed V-Agdal, Rabat	Examineur
Mlle. ECHABBI Loubna	Professeur Assistant, INPT, Rabat	Invitée
M. BELHADJ Abdenabi	Responsable Veille Technologique, Maroc Telecom	Invité



Laboratoire LIA, Avignon

Laboratoire LIMIARF, Rabat



Avant-propos

Cette thèse s'est déroulée au Laboratoire Informatique d'Avignon (LIA) de l'Université d'Avignon et des Pays de Vaucluse (UAPV) France conjointement avec le Laboratoire d'Informatique Mathématiques appliquées Intelligence Artificielle et Reconnaissance de Formes (LIMIARF) de la Faculté des Sciences de l'Université Mohammed V Agdal (UMVA), Rabat Maroc.

Cette thèse est réalisée en co-direction dans le cadre d'une cotutelle entre l'Université Mohammed V-Agdal Faculté des sciences, Rabat Maroc et l'Université d'Avignon et des pays de Vaucluse Avignon France. Elle est financée dans le cadre d'un projet porteur R&D de Maroc Telecom sous le contrat N°10510005458.06 PI.

Mes premiers remerciements iront à mon directeur de thèse le professeur El-Houssine Bouyakhf, pour m'avoir soutenu durant ma formation du DESA ITM et mes trois années de thèse. J'aimerais lui adresser mes plus vifs remerciements pour tout son dynamisme, son soutien, ses conseils et sa confiance qui m'ont permis de mener à bien cette thèse. Je souhaite lui transmettre l'expression de ma reconnaissance et ma plus profonde gratitude. Je tiens à remercier monsieur Eitan ALMAN Directeur de recherche à l'INRIA Sophia Anti-Polis Nice France pour ces qualités humaines, son rigueur et sa grande culture scientifique, qui ma donné l'occasion de faire ma thèse sous sa direction au sein du laboratoire LIA. Je voudrais également remercier Mr. Rachid El-Azouzi mon co-directeur de thèse à l'université d'Avignon, sans son encadrement exemplaire et complet, ce travail n'aurait jamais pu aboutir. Mr Rachid El-Azouzi qui m'a toujours consacré une grande partie de son précieux temps, m'a guidé et m'a conseillé. Ses directives pertinentes et ses compétences scientifiques sont autant d'éléments qui ont favorisé le développement de cette thèse. Je lui suis grandement reconnaissant et souhaite lui transmettre l'expression de ma reconnaissance.

Je remercie tout particulièrement les membres de mon jury de thèse présidé par le Professeur El Houssine Bouyakhf directeur de l'UFR ACSYS, qui ont accepté de juger ce travail et de participer au jury. Je tiens à remercier le Professeur Mohammed Benkhalifa de la faculté des sciences de Rabat, d'avoir accepté de faire partie du jury. Je remercie également Monsieur Rachid Elkouch, Professeur à l'INPT (Institut National des Postes et Télécommunications) et Monsieur Jalel Ben-Outhman, Maître de conférence HDR à l'Université de Versailles Saint-Quentin-en-Yvelines, France d'avoir bien voulu rapporter ma thèse. Je tiens à remercier également Mademoiselle Loubna Echabbi professeur assistant à l'INPT de s'intéresser à ce travail et d'être dans le jury. Enfin, je remercie

Monsieur Abdenabi BELHADJ responsable de la Veille Technologique du département Recherche et Développement (R&D) à Maroc Telecom, d'avoir accepté d'être parmi les membres de jury de ma thèse.

J'adresse toute ma gratitude et reconnaissance à ceux qui ont contribué à la réalisation, la correction, l'amélioration ainsi que la relecture du manuscrit.

Quoique je fasse, je ne peux remercier assez tous les membres des Laboratoires LIA et LIMIARF. En particulier, je pense au professeur Frédéric Béchet au LIA, avec qui j'ai partagé le bureau tout au long des trois années de thèse, merci pour ces qualités humaines, scientifiques et précieux conseils.

Je ne saurais jamais remercier ma famille : Mes parents, mes soeurs, mes frères, toi aussi à mon oncle, je vous dis "Merci". Plus particulièrement, mes remerciements vont aux êtres qui me sont les plus chers, qui ont eu un rôle essentiel et continu pendant mon cursus d'études, et qui sans eux aucune réussite n'aurait été possible. J'adresse de tout mon coeur mes remerciements à ma chère mère qui fut toujours mon seul exemple, je lui suis infiniment reconnaissant pour son amour, son soutien inconditionné malgré la distance. Et plus que ça ses encouragements à être toujours à la hauteur des responsabilités. Qu'elle trouve dans ce travail le fruit de ses efforts. Aussi, le fruit de ce travail ira à la mémoire de mon père qui m'a soutenu durant toute sa vie. Un soutien permanent tout au long de mes études, sans lequel je ne serai jamais arrivé à ce stade de réussite.

Resumé

Les systèmes de communications mobiles ont considérablement évolué durant ces dernières années. Ce fait a encouragé le déploiement de plusieurs systèmes ou réseaux cellulaires multi technologique. La qualité de service (QoS) offerte aux utilisateurs mobiles s'améliore d'un système à l'autre. Les systèmes de troisième génération (3G), comme l'UMTS, offrent une meilleure qualité de service par rapport à celle offerte par ceux de deuxième génération (2G), comme le GSM. Pour exemple, les systèmes de 3.5G (HSDPA) améliorent le débit du réseau de 3G sur le lien descendant afin de répondre aux exigences des nouveaux services. De plus, les réseaux de quatrième génération (4G), comme le WiMAX (IEEE802.16e) permettent, quant à eux, d'élargir la couverture cellulaire tout en offrant un débit supérieur. Cette nouvelle évolution améliore encore l'accessibilité aux services de l'Internet. La migration des services de l'Internet vers les réseaux mobiles, constitue un enjeu majeur de recherche en télécommunications.

La thèse se divise en deux objectifs principaux ; le premier consiste à calculer et à évaluer les performances du système UMTS-WCDMA. Deux approches sont utilisées, l'approche exacte de l'analyse spectrale, et l'approche approximative de la perturbation singulière. Le deuxième objectif propose de nouveaux mécanismes de contrôle d'admission prenant en considération la mobilité intra et inter cellulaires des utilisateurs dans la gestion de la ressource radio. Ces mécanismes d'allocation de ressource concernent plus particulièrement les réseaux HSDPA et WiMAX. Ils se basent principalement sur les schémas de modulation et de codage adaptatif AMC (*Adaptive Modulation and Coding*) : la zone géographique est subdivisée en plusieurs régions concentriques. Avec nos mécanismes d'allocation des ressources, chaque appel accepté par le système voit son débit initial maintenu, et ce, indépendamment de la position de l'utilisateur dans la cellule.

Mots clés : Réseaux sans fils, Evaluation de performance, Allocation de la ressource radio, Mobilité, Contrôle d'admission.

Abstract

The mobile communication systems evolved considerably in the last few years. This fact encouraged the deployment of several systems or cellular networks in multi technological environment. The Quality of Service (QoS) offered to the mobile users improves from one system to another one. The systems of third generation (3G), like UMTS, offer a better quality of service compared to that offered by those of second generation (2G), like the GSM. For example, the systems of 3.5G (HSDPA) improve the throughput of the network of 3G in the downlink direction according to the requirements of the new services. Moreover, the networks of fourth generation (4G), like WiMax (IEEE802.16e), as for them, make it possible to widen the cover of the base station while offering a very important throughput in which supports the next generation applications or services already offered by the Internet. This new evolution still improves accessibility with the services of the Internet. The migration of the services of the Internet towards the mobile networks, constitutes a major stake of research in telecommunications.

The thesis is divided into two main aims. The first consists to compute the system capacity and to evaluate the performances of the UMTS-WCDMA system. Two approaches are used, one is of the spectral analysis, and other one is an *approximation* of the singular perturbation. The second goal of this thesis is to propose and develop the new mechanisms of call admission control (CAC) for HSDPA and IEEE802.16e networks based on the AMC (Adaptive Modulation and Coding) scheme : the geographical area of the cell is subdivided into several concentric regions. These CAC mechanisms maintain a same QoS (constant bit rate) for all users anywhere in the area of the cell and give priority to migrating or handoff calls by reserving the guard bandwidth than a new call in the intra cell and inter cell mobilities.

Key words : Wireless networks, Evaluation of performance, Resource allocation, Mobility, Call admission control.

Table des matières

0.1	Motivations	12
0.2	Objectifs et contributions	13
0.3	Organisation du manuscrit	15
I	Généralités et état de l'art	17
1	Généralités	19
1.1	Concepts généraux	19
1.1.1	Notions de base	19
1.1.2	Les systèmes de communications mobiles étudiés	25
1.1.3	Présentation de l'UMTS	26
1.1.4	Présentation de HSDPA	30
1.1.5	Présentation de IEEE 802.16e (WiMAX)	32
1.2	Travaux existants	35
II	Évaluation des performances dans le réseau UMTS/W-CDMA	41
2	Évaluation des performances par l'Analyse Spectrale	43
2.1	Introduction	43
2.2	Cas de deux classes de service	45
2.2.1	Formulation du problème	45
2.2.2	Cas d'une cellule WCDMA isolée	47
2.2.3	Analyse du système	49
2.2.4	Approche par l'analyse spectrale	51
2.2.5	Distribution stationnaire	56
2.2.6	Métriques de performances	58
2.2.7	Cas multi cellules	59
2.2.8	Résultats numériques	60
2.3	Cas de trois classes de service	65
2.3.1	Contrôle d'admission du trafic temps réel	65
2.3.2	Contrôle d'admission du trafic non temps réel	67
2.3.3	Trafic <i>best effort</i> sans contrôle d'admission	69
2.3.4	Évaluation des performances par l'analyse spectrale	71
2.3.5	Illustrations numériques	74

2.4	Conclusion	76
3	Évaluation des performances par la Perturbation Singulière	79
3.1	Introduction	79
3.2	Formulation du problème	81
3.2.1	Hypothèses	81
3.3	Série de Taylor	83
3.3.1	Définitions et Hypothèses	83
3.3.2	Représentation analytique de la distribution stationnaire	86
3.3.4	Temps de séjour d'appels <i>best effort</i>	88
3.4	Résultats numériques	89
3.5	Conclusion	93
III	Gestion des ressources des réseaux mobiles HSDPA et IEEE802.16e	95
4	Gestion des ressources du réseau HSDPA en présence de la mobilité	97
4.1	Introduction	97
4.2	Formulation du problème	99
4.2.1	Modulation et codage adaptatif (AMC)	99
4.2.2	Bande passante par type d'appel	100
4.2.3	Modélisation	102
4.3	Mobilité intra cellule	102
4.3.1	Sans mobilité	102
4.3.2	Avec intra mobilité	103
4.3.3	Calcul des taux de migration	105
4.4	Mobilité inter cellules	106
4.5	Métriques des performances	107
4.6	Résultats numériques	109
4.7	Conclusion	117
5	Gestion des ressources du réseau mobile IEEE 802.16e WiMAX	119
5.1	Introduction	119
5.2	Analyse du système IEEE 802.16e dans le cas discret	121
5.2.1	Description du système	121
5.2.2	Mécanisme de contrôle d'admission	123
5.2.3	Modèle	124
5.2.4	Allocation de ressources et Matrice de transition	126
5.2.5	Métriques des performances	129
5.2.6	Exemples numériques	131
5.3	Analyse du système IEEE 802.16e dans le cas continu	137
5.3.1	Description du système	137
5.3.2	Mécanismes de contrôle d'admission	137
5.3.3	Modèle	139
5.3.4	Comportement des utilisateurs mobiles	141
5.3.5	Analyse du système avec le premier mécanisme	143

5.3.6	Analyse du système avec le deuxième mécanisme	145
5.3.7	Evaluation des performances	147
5.3.8	Applications numériques	148
5.4	Conclusion	154
	Liste des publications	158
	Liste des illustrations	165
	Liste des tableaux	167
	Liste des indexes	169
	Bibliographies	171

0.1 Motivations

Les évolutions du monde des télécommunications vers le multimédia mobile suite aux avancées technologiques ont démontré que fournir l'accès au réseau n'est plus suffisant. Le besoin des utilisateurs s'oriente vers l'accès aux services multimédia à valeur ajoutée, dans leur propre environnement nominal indépendamment de leur mode d'accès aux systèmes. Les services multimédia nécessitent de hauts débits de transfert et ont des exigences en qualité de service. Ils doivent coexister avec les services à contraintes temps réel tel que le service voix qui ne tolère pas de variation du délai entre l'envoi et la réception de paquets. La garantie de ces services par l'opérateur devient beaucoup plus difficile dans les technologies qui prennent en considération la mobilité des utilisateurs.

L'accessibilité totale en toute mobilité à des services de plus en plus consommateurs en ressources réseau et en traitements interactifs et temps réels, nécessitent des solutions et des supports technologiques pour gérer les multiples modes d'accès, de transport et les contrôles spécifiques à chaque service ou application, comme le service temps réel et le transport fiable de données, la vidéo conférence ou le transfert de flux vidéo temps réel, comportement transactionnel, paiement, etc.

Des efforts ont notamment été déployés dans la perspective de gérer au mieux la qualité de service et également pour différencier les services temps réel et best effort ou non temps réel. Il s'agit d'offrir un service ainsi qu'un contenu soumis à des contraintes de disponibilité de ressources radio et de charge du réseau, tout en conservant la sémantique et en offrant un rendu conforme aux attentes des utilisateurs.

De nouveaux mécanismes s'imposent alors pour différencier les services et pour fournir la qualité de service exigée. Parmi ces mécanismes, on compte les contrôles d'admission, les protocoles de réservation de ressources et les politiques d'ordonnancement des paquets. La difficulté pour les réseaux de nouvelle génération est de maintenir les hauts débits et la qualité de service sur les liens radio (montant et descendant) qui sont peu fiables pour les transmissions. Le CDMA est la technique la plus aboutie du multiplexage, destinée à être utilisée notamment sur les réseaux de téléphonie mobile de troisième génération comme l'UMTS. Alors que les précédentes techniques de multiplexage en fréquence (FDMA) et de multiplexage temporel (TDMA) consistaient essentiellement à diviser une grandeur physique (une bande passante de fréquences ou le temps qui s'écoule) en " créneaux " individuels, le CDMA ne fixe pas de répartition statique préalable de ces ressources et s'apparente aux techniques de transmission informatique par paquets. Cette technique a été choisie en raison des contraintes propres à l'interface radio de 3G : débit variable, Qualité de service variable (environnement de trafic multiservices). Elle autorise l'accès simultané de multiples utilisateurs à de très haut débits et à des services très variés (son, image, multimédia, ...). Il existe néanmoins différentes variantes de CDMA. Celle sur laquelle s'appuie la norme UMTS appelée WCDMA. Le débit promis par l'UMTS en théorie est 2 Mbits/s, mais il n'est jamais atteint en pratique à cause des limites du réseau de 3G. Une extension logiciel de WCDMA appelée HSDPA a été introduite afin d'améliorer le débit du lien descen-

dant où plus d'informations seront transportées. Avec la technique HSDPA le débit peut atteindre 14 Mbits/s dans le réseau de 3.5G ou bien HSP+. Pour mieux gérer les ressources radio par l'opérateur avec plus de souplesse en terme de qualité de service et de débit, une nouvelle méthode d'accès multiple OFDMA est devenue la base des réseaux de nouvelle génération, comme IEEE80.16e et LTE. Le réseau WiMAX permet un débit théorique de 70 Mbits/s sur un rayon de 50 km maximum tout en se basant sur de nouvelles techniques, comme le schéma de modulations et du codage adaptatifs en fonction de l'état du canal radio associé à un mobile. LTE quant à lui promettra un très haut débit allant jusqu'à 100 Mbits/s ainsi qu'une large accessibilité de couverture (100 km en zone rurale).

0.2 Objectifs et contributions

Les objectifs tracés pour cette thèse sont de proposer des nouvelles méthodes pour l'évaluation des performances du système WCDMA de manière explicite, de comparer ces méthodes au niveau du temps de calcul du à la complexité de ces méthodes et de valider les formules obtenues en utilisant le simulateur NS-2, et enfin d'exposer des nouveaux mécanismes d'allocation des ressources par rapport à la mobilité des utilisateurs dans les réseaux haut débit HSDPA et WiMAX (IEEE802.16e). Les contributions apportées dans cette thèse sont les suivantes :

- Nous présentons une évaluation des performances du système WCDMA dans un premier scénario de deux classes de service : temps réel et *best effort*. Le trafic temps réel pour des flux temps réel de paquets de taille fixe émis périodiquement, mode particulièrement adapté aux conversations vocales qui bénéficient d'une bande passante contrôlée avec échange interactif en temps réel et avec un minimum de délai entre les paquets. Le trafic *best effort* est destiné pour un trafic qui ne demande aucune qualité de service, il est particulièrement adapté au mode interactif qui est plutôt destiné à des échanges de type navigation sur le Web (requête/réponse).

En particulier, nous obtenons via la méthode d'analyse spectrale les probabilités d'états d'équilibres des nombres d'appels BE en fonction des valeurs propres et des vecteurs propres de certaines matrices finies. Ce résultat nous permettra d'évaluer les performances du système en fonction des valeurs propres et vecteurs propres de certaines matrices. Toutefois, nous montrons que ces valeurs propres et vecteurs propres sont obtenus *indépendamment* du nombre d'appels BE qui réduit la complexité de calcul de la distribution stationnaire. Par conséquent, notre approche est utile pour un opérateur dans sa politique d'admission d'appels *best effort*, avec moins de calcul, qui maximise l'utilisation de la bande passante et garantit la QoS exigée par les appels temps réel et *best effort*.

Dans un deuxième scénario, nous introduisons une troisième classe appelée "non temps réel" pour des flux demandant des débits minimums, par exemple comme le transport des fichiers FTP. Cette classe est intermédiaire entre la classe temps réel et la classe *best effort* citées ci-dessus. Nous présentons une analyse du système WCDMA avec une chaîne de Markov par la même méthode d'analyse spec-

trale. Nous signalons que cette méthode donne une solution exacte de la distribution stationnaire dans le cas de trois classes de service. Mais, le temps de calcul devient considérable et très important pour obtenir des performances du système. Nous introduisons un nouveau mécanisme du CAC pour les appels non temps réel en nous basant sur leur temps de séjour qui doit être borné. Le nouvel appel non temps réel est accepté si le délai correspondant au nombre actuel d'appels non temps réel est borné par une constante positive fixée par l'opérateur. Le mécanisme du CAC gère un nombre maximum d'appels non temps réel pour assurer leurs délais. Il permet à la classe *best effort* d'utiliser une partie de la bande passante libérée par les appels non temps réel, voir le chapitre 2 ;

- Nous constatons que les transitions en terme, du nombre d'appels *best effort* varient très lentement et celles du nombre d'appels temps réel varient très rapidement. Ce qui nous permet de définir deux échelles de temps : une rapide et l'autre lente. Dans ces conditions particulières, nous utilisons une approximation simple de la distribution stationnaire appelée perturbation singulière pour évaluer les performances du système WCDMA. Nous obtenons la distribution d'états d'équilibre du système WCDMA de manière approximative avec moins de calcul qu'auparavant et proposons donc un contrôle d'admission moins complexe. Les performances du système sont obtenues indépendamment du nombre d'appels *best effort* et sont décrites dans le chapitre 3. Contrairement à celles trouvées dans le chapitre 2 qui dépendent directement de ceux en temps réel. Dans ce cas, quand le nombre d'appels temps réel augmente, le système prendra un temps important pour le calcul des performances. Et s'il y a des erreurs dû aux interférences des évanouissements, d'effet de masque, etc., le système doit refaire tout le calcul mathématique. Ce qui rendra le système plus lourd au niveau temps de calcul, d'où l'avantage de cette approximation ;
- Nous nous intéressons dans ce chapitre à l'étude de la qualité de service dans le réseau haut débit HSDPA extension du réseau UMTS-WCDMA. La qualité de service dépend directement de la qualité du canal radio associée au mobile et varie selon son état (bon ou mauvais). Cette qualité de service devient trop compliquée à garantir et varie selon le niveau de la mobilité faible ou forte, et ainsi, selon le type de mobilité intra ou inter cellule de l'utilisateur. Nous développons un nouveau mécanisme de contrôle d'admission qui permettra au mobile d'avoir son débit initial durant son mouvement dans la cellule/le réseau. Nous définissons deux types de mobilité intra et inter cellule. La mobilité intra cellule correspond aux changements de modulation (état du canal) au sein de la même station de base. Alors que, la mobilité inter cellule se définit comme le changement de la station de base pour un mobile dans le réseau. Toutefois, nous décomposons la cellule HSDPA en un nombre fini de régions de même centre et chacune d'elles est associée à une modulation donnée (état moyen du canal de transmission). Nous calculons explicitement le taux de migration d'appels migrant d'une région à une autre en fonction du nombre moyen estimé d'appels actifs dans cette région et de la probabilité de migration. En plus, le nouveau mécanisme affectera de manière dynamique une bande passante nécessaire pour maintenir le débit initial d'un appel (temps réel ou non temps réel) quelque soit sa position dans la cellule.

Ce mécanisme donne plus de priorité aux appels migrant en leurs réservant une bande passante uniquement pour eux. La valeur de cette bande peut être contrôlée par l'opérateur selon les périodes de mobilité (forte ou faible) du réseau, voir le chapitre 4 ;

- Nous introduisons également des nouveaux mécanismes de contrôle d'admission dans le réseau IEEE802.16e (WiMAX) avec deux types de trafics temps réel et non temps réel. Deux cas de figures sont étudiés : cas discret et cas continu du système IEEE 802.16e multiservices. Dans le premier cas, nous nous intéressons uniquement à la qualité de service d'appels temps réels en leurs garantissant un débit quelque soit la position du mobile dans la cellule. Pour ce faire, nous décomposons la cellule IEEE802.16e isolée selon la technique de modulation et de codage adaptatif AMC (*Adaptive Modulation and Coding*) en plusieurs régions. Une approche basée sur une chaîne de Markov à temps discret est développée prenant en compte les ressources libres pour bloquer ou accepter un nouvel appel temps réel. Par contre, les appels non temps réel partagent équitablement les ressources restante selon le processus *Sharing*. Dans ce premier cas nous ne traitons pas la mobilité des utilisateurs dans la gestion de la ressource radio.

Dans le deuxième cas, nous visons le cas continu du système IEEE802.16e. Nous développons deux nouveaux mécanismes d'allocation de ressources prenant en compte cette fois ci, la mobilité intra cellule des utilisateurs. Le premier mécanisme permettra aux appels temps réel et non temps réel d'obtenir un même débit moyen tout le temps durant la connexion du mobile. Dans ce cas, une bande passante est réservée pour gérer le problème de la mobilité pour les deux classes d'appels. Le deuxième mécanisme quant à lui, est conçu pour introduire une qualité de service minimale pour les appels non temps réel en leurs réservant une bande passant minimale et garde la même qualité de service pour les autres appels. Il gère également la mobilité des deux classes d'appels en leurs réservant une bande passante. Dans ces deux mécanismes d'allocation de ressources, nous considérons un modèle de mobilité appelé RWP (*Random Waypoint*), voir le chapitre 5.

0.3 Organisation du manuscrit

Nous avons organisé ce manuscrit en trois parties. La première partie est constituée du chapitre 1 sur les concepts généraux en relation avec cette thèse et de l'état de l'art. La deuxième partie est constituée des chapitres 2 et 3. Elle concerne l'évaluation des performances du système WCDMA en lien montant. La troisième partie, composée des chapitres 4 et 5, est dédiée à la gestion des ressources radio en lien descendant des réseaux HSDPA et WiMAX. Cette organisation est la suivante :

- Le chapitre 1 présente quelques concepts généraux qui seront utilisées tout au long de cette thèse à savoir les méthodes d'accès multiples radio de base, le codage et la modulation adaptative, l'évolution des systèmes de communication sans fil, l'architecture du réseau UMTS et les services qu'il offre, le système HSDPA,

-
- le réseau IEEE802.16e (WiMAX) et les services qu'il offre, ainsi que les travaux antérieurs ;
- Le chapitre 2 expose les performances du lien montant (*Uplink*) du système WCDMA. L'étude est faite d'abord pour deux type de trafics temps réel et *best effort*. En générale, les performances du système WCDMA dans la littérature se basent sur des solutions numériques, voir par exemple (Hegde et Altman, 2003). Les métriques de performances dans ce chapitre sont obtenues explicitement en fonction des paramètres du système à savoir le taux d'arrivée et le taux de service d'appels ainsi que le nombre d'appels. Un nouveau trafic non temps réel intermédiaire entre ces deux trafics est introduit. Egalement, les performances du système sont obtenues explicitement en nous basant sur l'approche d'analyse spectrale. Le chapitre suivant présente une autre solution alternative.
 - Le chapitre 3 présente une solution approximative de la distribution stationnaire des nombres d'appels temps réel et *best effort* pour évaluer les performances du système WCDMA. Avec quelques hypothèses à respecter, le système peut être analysé par l'approche d'approximation appelée perturbation singulière. La distribution stationnaire du système WCDMA est obtenue sous forme d'une série de Taylor selon le niveau d'approximation. Avec cette méthode, moins de calcul sera fait par un contrôle d'admission ;
 - Le chapitre 4 présente une stratégie d'allocation de ressource dans le réseau à haut débit HSDPA. Un contrôle d'admission est proposé. Il divise la surface de couverture d'une station de base (Node-B) en plusieurs régions selon le principe d'efficacité de modulation AMC. Le mécanisme du CAC fait distinguer deux trafics temps réel et non temps réel en nous basant sur le type de service demandé par l'utilisateur. Il affecte de manière dynamique une bande passante effective à chaque appel accepté dans le système en fonction de son efficacité de modulation et garde son débit initial durant sa communication ;
 - Le chapitre 5 étudie le système IEEE802.16e selon deux cas de modélisation : discret et continu. Dans les deux cas, un modèle de la cellule IEEE802.16e est proposé et permet de la décomposition de la cellule selon le principe de la technique du codage et modulation adaptatif AMC. Le modèle se base sur un mécanisme de contrôle d'admission en présence de deux types de trafics temps réel et non temps réel. Dans le premier cas, le modèle basé sur un nouveau mécanisme du CAC garantit un même débit aux appels temps réel et aucune QoS aux appels non temps réel. Le CAC ne prend pas en considération la mobilité des utilisateurs intra cellule. Dans le deuxième cas, le modèle se base sur deux nouvelles stratégies du CAC avec cette fois-ci la mobilité intra cellule et donne une même QoS pour les appels de ces deux trafics en favorisant les appels en cours sur les nouvelles arrivées ;
 - Enfin le chapitre 5.4 conclut cette thèse en établissant un bilan sur les travaux effectués et des perspectives de travaux futur.

Première partie

Généralités et état de l'art

Chapitre 1

Généralités

Sommaire

1.1 Concepts généraux	19
1.1.1 Notions de base	19
1.1.2 Les systèmes de communications mobiles étudiés	25
1.1.3 Présentation de l'UMTS	26
1.1.4 Présentation de HSDPA	30
1.1.5 Présentation de IEEE 802.16e (WiMAX)	32
1.2 Travaux existants	35

1.1 Concepts généraux

L'objectif de ce chapitre est de présenter quelques notions de base, des systèmes de communications mobiles et les travaux antérieurs qui ont une relation avec cette thèse. Nous y présentons, dans cette première sous section (1.1.1), quelques types de techniques d'accès multiples, le codage et la modulation adaptatif, les systèmes de communication mobile étudiés dans cette thèse (UMTS, HSDPA et WiMAX). Enfin, dans la deuxième sous section (1.2), nous terminons par la présentation de quelques travaux contextuels déjà existant dans la littérature.

1.1.1 Notions de base

Dans cette section, nous allons présenter quelques méthodes d'accès radio de base (Smith et Collins, 2002).

FDMA (*Frequency Division Multiple Access*) : La technique d'accès multiple à division fréquentielle, FDMA, représente la technique la plus utilisée dans les réseaux mobiles de première génération. Avec cette méthode, le spectre est divisé en plusieurs

canaux radio (aussi appelées voies ou bandes incidentes, sous-bandes de fréquences) avec une largeur de bande fixée, voir la figure 1.1 (a). Quand un utilisateur désire effectuer un appel, un des canaux est alors exclusivement alloué à cet utilisateur pendant toute la durée de la communication. Un canal de garde se situe entre chacun de ces canaux pour éviter l'intermodulation. Cette technique s'applique principalement aux signaux analogiques.

TDMA (*Time Division Multiple Access*) : La technique d'accès multiple à division de temps, TDMA, est surtout utilisée dans les systèmes de deuxième génération, tels que le GSM. Par cette technique, une même bande de fréquences est partagée par un certain nombre d'utilisateurs qui se voient attribués un intervalle (appelé *slot* ou *IT*) de temps unique, voir la figure 1.1 (b). L'ensemble des IT mis ensemble forment une trame. Dans le cas de la norme européenne GSM, chaque canal de 200 kHz est divisé en huit tranches de temps, tandis que selon la norme américaine IS-136, chaque canal de 30 kHz est divisé en trois tranches de temps. Ces systèmes permettent d'avoir, dans des conditions réelles d'utilisation, une capacité en terme de nombre d'utilisateurs 3 à 6 fois supérieure à celle des systèmes qui se basent sur le FDMA.

CDMA (*Code Division Multiple Access*) : La technique CDMA constitue une méthode d'accès où tous les utilisateurs partagent la même bande de fréquences simultanément et tout le temps, comme le montre la figure 1.1 (c). La technique d'étalement de spectre est utilisée pour affecter à chaque utilisateur un code, ou séquence, unique qui détermine les fréquences et la puissance utilisées. Le signal contenant l'information de l'émetteur est modulé avec la séquence qui lui est attribuée, ensuite le récepteur recherche la séquence en question. En isolant toutes les séquences provenant des autres utilisateurs (qui apparaissent comme du bruit), le signal original de l'utilisateur peut alors être extrait.

W-CDMA (*Wideband-Code Division Multiple Access*) : Il s'agit de la même technique que CDMA, en utilisant des canaux large de 5 Mhz en UMTS. Le protocole WCDMA se base principalement sur le processus d'étalement du spectre à séquence directe DS-SS-CDMA (*Direct Sequence-CDMA*). Dans ce type d'étalement de spectre, le signal d'information est directement modulé par une séquence ou code appelé « code d'étalement ». La figure 1.2 illustre le fonctionnement de cette technique ; la figure 1.2 a) montre les principaux éléments d'un émetteur DS-SS-CDMA. Parmi ces éléments, on trouve principalement un générateur de codes d'étalement et un « multiplicateur ». Le code généré constitue une séquence de longueur fixe dont chaque élément constitue un « chip ». Le processus d'étalement est illustré dans la figure 1.2 b), qui montre le code d'étalement du k -ième utilisateur qui est représenté par la séquence de chips $c_k^{(p)}$, où $p = 1, 2, 3, \dots, M$ et où M représente la longueur du code ($M = 4$ dans cet exemple). Chaque chip a une durée T_c et une amplitude égale à +1 ou -1. Le débit avec lequel varie l'information étalée est appelé débit *chip* (ou *chip rate*) que l'on notera B_{spr} . Ce débit est égal à $1/T_c$, exprimé en chips par seconde (cps).

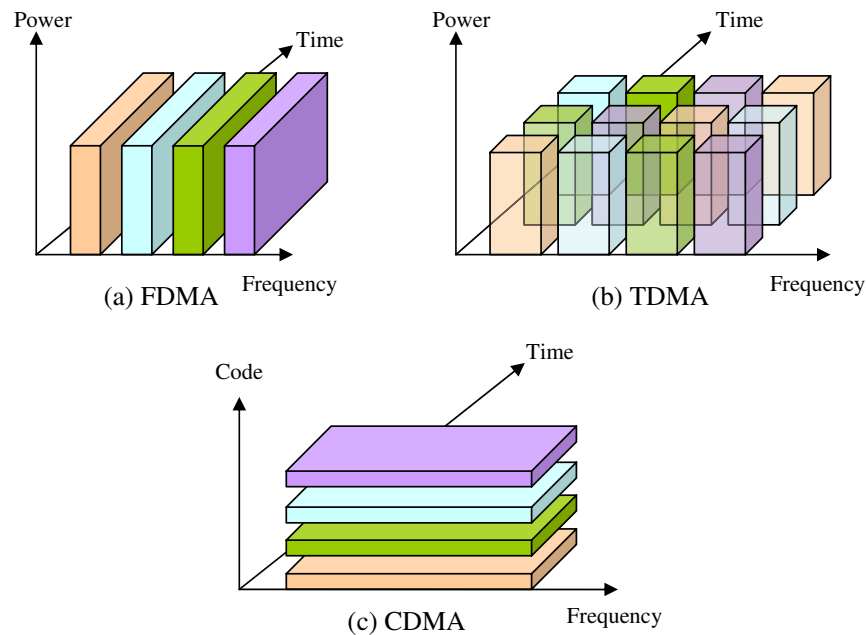


FIGURE 1.1 – Méthodes d'accès multiples (Alwis, 2005).

D'autre part, le n -ième symbole du message d'information de l'utilisateur k est noté $d_k^{(n)}$, où $n = 1, 2, \dots$. Ainsi, et de manière similaire, nous pouvons aussi noter la durée de chaque symbole par T_s , à partir de laquelle nous pouvons également formuler le débit symbole $B_s = 1/T_s$ exprimé en symboles par seconde (sps). Il est important de constater que ce débit symbole est variable dans un système DS-CDMA, car il dépend du service offert (voix, vidéo ou données), tandis que le débit *chip* B_{spr} est constant et fixé à 3.84 Mcps (Millions de chips par seconde).

Le rapport entre le débit du signal étalé et le débit du signal non étalé est appelé facteur d'étalement SF (*Spreading Factor*). Il représente une propriété importante de l'étalement de spectre et s'exprime par : $SF = \frac{B_{spr}}{B_s} = \frac{1/T_c}{1/T_s} = \frac{T_s}{T_c}$

Par conséquent, le facteur d'étalement dans un système DS-CDMA est égal au nombre de chips utilisés pour étaler un symbole d'information. Après la génération des codes, le processus d'étalement se déroule simplement en multipliant le message par +1 ou -1. Selon la figure 1.2 b), il apparaît clairement que $T_s > T_c$, d'où $B_{spr} > B_s$. On comprend alors comment on passe d'un signal à bande étroite vers un signal à large bande. Finalement, le signal subit une modulation en fréquence et il est mis sur une fréquence porteuse avant d'être envoyé. Au niveau de la réception, il suffit de reprendre la chaîne dans l'ordre inverse, notamment en démodulant dans un premier temps le signal reçu, pour ensuite le multiplier par le même code d'étalement utilisé lors de l'émission. À la suite de cette multiplication, nous obtiendrons le message d'information « désétalé » (exprimé en débit symbole).

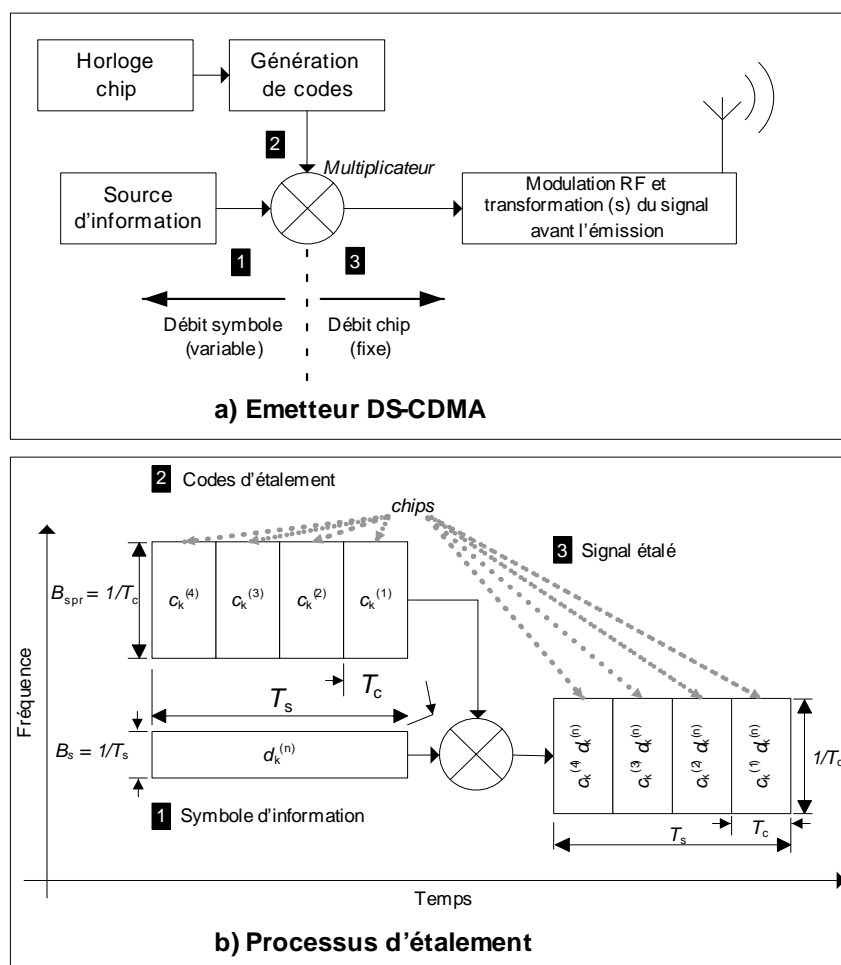


FIGURE 1.2 – Étalement à séquence directe DS-SS-CDMA.

OFDMA (Orthogonal Frequency Division Multiple Access) : Dans la technique OFDMA, l'ensemble des sous porteuses d'un symbole OFDM (*Orthogonal Frequency Division Multiple*) (Andrews, 2007) est divisé en sous ensembles de fréquences, ces derniers peuvent être affectés à différents utilisateurs. L'OFDMA apporte une nouvelle dimension lors de l'allocation des ressources : il faut affecter les différentes fréquences aux utilisateurs pendant un même *slot* comme le montre la figure 1.3. Durant un même *slot*, plusieurs utilisateurs peuvent occuper des fréquences différentes et ces mêmes fréquences peuvent être assignées aux mêmes utilisateurs durant les prochains *slots* en fonction de leurs besoins. Ces fréquences sont espacées par les canaux de garde ce qui annule les interférences entre les utilisateurs ou intra cellule. Pour des applications fixes ou nomades où les canaux ne varient que faiblement, l'avantage indiscutable de l'OFDMA sur les autres méthodes à accès multiple est sa capacité d'exploiter la niche des terminaux utilisateurs embarqués. La technique OFDMA est la base du réseau sans fil IEEE802.16e (WiMAX) comme défini dans le standard (IEEE Std 802.16e, 2005).

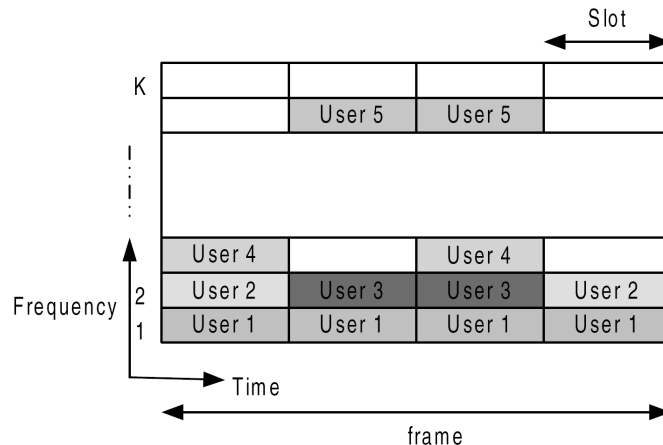


FIGURE 1.3 – Trame OFDMA.

Codages et modulations adaptatifs

Dans les réseaux mobiles, la qualité d'un signal reçu par un terminal dépend de plusieurs facteurs, comme la distance entre la station de base ou la perte de propagation due aux obstacles ou aux trajets multiples. Dans le but d'améliorer la performance d'un système, en termes de débit de données et de fiabilité de la couverture radio, le signal transmis vers et par un utilisateur particulier est modifié de façon à prendre en compte les variations de la qualité du signal à travers un processus connu sous le nom d'«adaptation au lien radio» (*link adaptation*), connu également sous le nom de «AMC» (*Adaptive Modulation and Coding*).

Une des techniques majeures introduites dans le WCDMA est le contrôle de puissance. L'idée est d'augmenter la puissance de transmission quand la qualité du signal reçu est faible et de la diminuer quand la qualité du signal en réception atteint un seuil donné. Ceci permet une communication fiable entre l'émetteur et le récepteur. Ainsi la technique de contrôle de puissance réduit les interférences intra et inter-cellules causées par une puissance de transmission trop importante, la capacité du système est ainsi augmentée. Une technique alternative au contrôle de puissance en faisant face aux effets variables du canal dans le temps est de compenser les évanouissements. Au lieu de garder une qualité de signal constante au niveau du récepteur, on peut changer la modulation et le code du signal transmis de telle façon que le plus d'informations soit transmis lorsque l'état du canal est bon et le moins possible lorsque le canal est détérioré. Cette technique est l'AMC. Comparé à la technique conventionnelle de contrôle de puissance, l'AMC apporte une capacité beaucoup plus élevée pour les systèmes radio à transmission de paquets comme HSPA (*High Speed Packet Access*) et WiMAX (*Worldwide Interoperability for Microwave Access*). Un autre avantage de l'AMC est que la puissance d'émission est fixe pendant toute la durée de la trame, ainsi les interférences provenant des autres utilisateurs sont significativement réduites. En plus de la modulation de base de la technique WCDMA définie dans la *Release 99* QPSK, des nouvelles modulations sont proposées dans les *Release 5, 6 et 7* (Holma et al., 2007), la modulation

16-QAM et la modulation 64-QAM pour la technique HSPA.

Mais, contrairement à la modulation QPSK, celles-ci sont optionnelles et dépendent du type de terminal utilisé. La modulation 16-QAM soutient 4 bits/symbole au lieu des 2 bits/symbole de la modulation QPSK. Ainsi, la modulation 64-QAM soutient 6 bits/symbole au lieu des 4 bits/symbole. Les débits sont ainsi augmentés de manière significative. La modulation 16QAM requiert de bonnes conditions de canal et une bonne performance des récepteurs comparée à la modulation QPSK. La décision d'une transmission en 64-QAM ou 16-QAM ou QPSK est faite dans le réseau en utilisant la qualité du canal d'information provenant du mobile via un canal de contrôle montant. En effet, les mobiles HSDPA effectuent des mesures de la qualité du canal physique descendant et transmet un indicateur de qualité de canal CQI (*Channel Quality Indicator*) dans le canal de contrôle montant HS-SCCH. Il faut noter que l'introduction d'un ordre de modulation plus élevé est accompagnée d'une plus grande complexité dans les terminaux mobiles, qui doivent estimer l'amplitude relative des symboles reçus. Cette estimation d'amplitude est nécessaire pour que le terminal puisse séparer tous les points de la constellation 16-QAM et 64-QAM telles qu'illustrés par la figure 1.4. Étant donné que cette estimation devient plus difficile lorsque la qualité du signal reçu est mauvaise, il est alors plus judicieux d'utiliser la modulation la plus faible dont la constellation est moins dense.

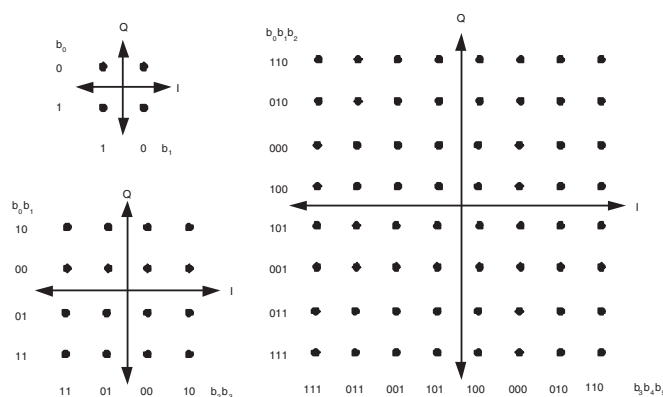


FIGURE 1.4 – Points de constellation pour QPSK, 16-QAM et 64-QAM.

En combinant le type de modulation, le taux de codage et le nombre de codes supportés, nous obtenons plusieurs combinaisons, appelées également schéma de modulation et de codage MCS (*Modulation and Coding Scheme*), qui ont été proposées par le 3GPP dans la Release 5. Ainsi, un utilisateur qui serait proche d'une station de base pourrait se voir attribué, par exemple, une modulation 64-QAM avec un taux de codage de 3/4 et dont le terminal supporte 15 codes en parallèles. Cette combinaison lui permettrait de bénéficier d'un débit plus élevé sur le lien descendant. L'AMC a été intégré aux fonctionnalités des Node B supportant la technologie HSDPA. Par conséquent, la station de base a la responsabilité de sélectionner l'algorithme de modulation et de codage approprié. Selon le nombre de codes parallèles pouvant être supporté par le terminal, nous obtenons plusieurs classes de terminaux HSPA spécifiées dans la Release 5 avec un to-

tal de 12 catégories. Ces catégories permettent d'avoir des débits pics allant de 0.9 à 14.4 Mbits/s. En outre, ce même principe de schéma de modulation a été adopté par le standard du réseau WiMAX (IEEE802.16e), comme le montre le tableau 1.1.

Modulation	Taux de codage	Bits/symbole	SNR (dB)
64-QAM	3/4	4.5	24.4
64-QAM	1/2	4	22.7
16QAM	3/4	3	18.2
16QAM	1/2	2	16.4
QPSK	3/4	1.5	11.2
QPSK	1/2	1	9.4
BPSK	1/2	0.5	6.4

TABLE 1.1 – Modulations et taux de codages adaptatifs sur le IEEE802.16e

1.1.2 Les systèmes de communications mobiles étudiés

Évolution des standards de communication

Depuis les premières expériences de Marconi à la fin du XIXe siècle, les communications sans fil entre deux éléments mobiles ont été limitées à des applications professionnelles et ce, jusqu'à la fin des années soixante dix. Il a fallu attendre les années quatre vingt avant de voir la naissance des systèmes de téléphonie mobile cellulaire, de première génération, entièrement analogiques comme le NMT (*Nordic Mobile Telephony*), l'AMPS (*Advanced Mobile Phone Service*), le TACS (*Total Access Cellular System*) ou le *Radiocom2000*. Néanmoins, les ressources spectrales réduites, la courte durée de vie des batteries et le coût du terminal, ont été des obstacles qui ont restreint les débuts du développement des téléphones portables de première génération. La seconde génération des systèmes cellulaires est apparue au début des années 90 avec le DECT (*Digital Enhanced Cordless Telecommunications*), le GSM (*Global System for Mobile communications*, Europe), le PDC (Japon), et le PCS (États-Unis). Prévus initialement pour des applications de transport de la parole et de données à faibles débits (9600 bits/s) autour des fréquences (935-960 MHz) ou (890-915 MHz), cette norme a permis l'essor des communications mobiles. Les performances (efficacité spectrale, sécurité, fonctionnalités) de la seconde génération de systèmes cellulaires sont plus importantes que celles de la première génération. Un important apport à ces systèmes a consisté en la modification de la méthode d'accès aux données, et a donné lieu à la naissance de la génération intermédiaire nommée 2.5G, incarnée par le système GPRS (*General Packet Radio Service*). Une autre forme d'évolution des systèmes cellulaires de seconde génération est incarnée par les systèmes EDGE (*Enhanced Data rate for GSM Evolution*); ils utilisent une nouvelle modulation, optimisant le débit efficace à l'interface GSM, permettant d'atteindre des débits de 384 Kbits/s au maximum. Les différents continents ont développé des systèmes assez similaires dans le principe, mais parfois différents dans leur réalisation. C'est donc dans le cadre d'une homogénéisation globale des standards qu'est née l'idée

d'une troisième génération de systèmes cellulaires. La troisième génération (3G) de systèmes cellulaires, souvent assimilée en Europe à l'UMTS (*Universal Mobile Telecommunication Services*), a été principalement conçue et imaginée afin de satisfaire l'acheminement de données multimédia (accès à Internet, aux réseaux d'entreprise, à des services de messageries, de visioconférence, de jeux, ...). Au même titre que le GPRS et EDGE ont permis, sans modifier le principe de base des communications, à leur génération de systèmes cellulaires d'évoluer en terme d'efficacité. Le HSDPA (*High-Speed Downlink Packet Access*) est déjà considéré comme le standard 3.5G des télécommunications mobiles. Alors que le débit maximum permis sur une liaison UMTS est de 2 Mbits/s pour une bande passante de 5MHz, le HSDPA, grâce à sa modulation en 16-QAM, autorise des débits de 10 Mbits/s. Ainsi, grâce à l'introduction de la nouvelle modulation 64-QAM dans la *Release 7*, le débit théorique de 14 Mbits/s est possible par le réseau HSDPA. Dans ce même sens, le réseau IEEE802.16 (WiMAX) dit 4G permet un débit théorique de 70 Mbits/s sur un rayon de 50 km maximum tout en se basant sur des modulations différentes en fonction de l'état du canal radio et de nouvelles méthodes d'accès au canal. Les performances du système en termes du débit sont beaucoup plus importantes que celles fournies par les systèmes antécédents, grâce à ces modulations (QPSK, 16-QAM, 64-QAM). La norme WiMAX Mobile permet des services comme la communication en VoIP (Téléphonie sur Réseau IP) sur téléphone portable ou encore l'accès à des services mobiles en hauts débits. Son équivalent en Europe est le HiperMAN. De plus, le but principal des nouvelles technologies de communication est de supprimer les câbles et de permettre la transmission (voix, données, ...) via une liaison radio. La figure 1.5 résume l'évolution des réseaux sans fil.

Ainsi, le développement rapide des communications sans fil et l'émergence des nouveaux standards sollicitent la convergence vers la quatrième génération de communications mobiles. Initialement prévue pour les années 2010, elle semblerait présenter ses toutes premières réalisations commerciales légèrement plus tôt. Cette avance par rapport aux prévisions est notamment due au fait que la quatrième génération ne sera pas l'issue d'une révolution dans les communications (comme c'est le cas de la deuxième et la troisième génération), mais plutôt la convergence de différents standards, applications et produits. En effet, il semble pertinent de considérer la 4G comme la convergence des standards et des technologies couverts par la 3G et les réseaux locaux sans fil (WLAN). Le but de la 4G est d'améliorer les performances de la troisième génération, sans changer fondamentalement le contenu ni les applications prévues au départ pour la 3G.

1.1.3 Présentation de l'UMTS

L'UMTS est un réseau mobile de troisième génération capable d'offrir des bénéfices significatifs à l'utilisateur en terme de services à valeur ajoutée, tels que l'accès Internet à haute vitesse, le téléchargement de fichiers (audio et vidéo) ou alors la visioconférence. L'UMTS se base principalement sur la technique d'accès multiple large bande WCDMA pour y offrir ce type de service. Le système universel UMTS a été choisi dans le but de faire une distinction avec les systèmes de première et de deuxième génération qui sont

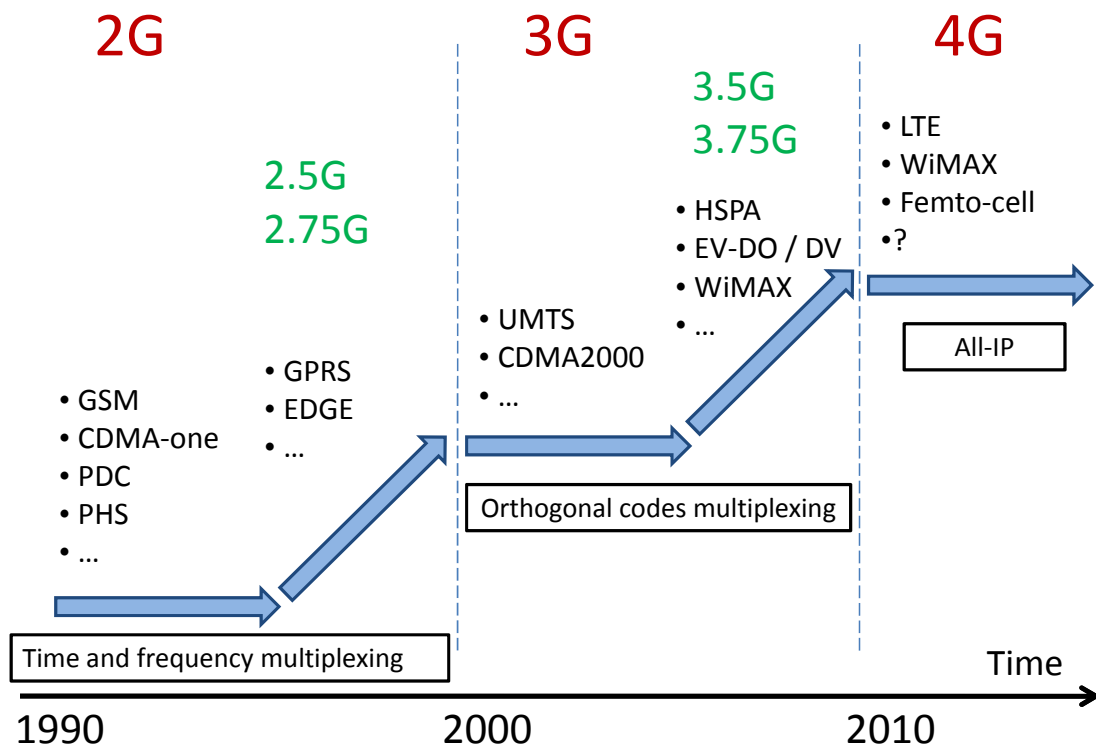


FIGURE 1.5 – Evolution des réseaux sans fil.

considérés comme des systèmes axés principalement sur le service de la voix. Dans cette section, nous allons présenter les catégories de services pouvant être offerts par un réseau UMTS. Puis, nous verrons son architecture en général. Nous verrons également les couches de protocoles radio de l'interface d'accès ainsi que la description des canaux de transport. Le 3GPP a rendu obligatoire l'utilisation du codeur de parole AMR (*Adaptative Multi Rate*) (Holma et Toskala, 2001) dans le réseau UMTS - alors qu'il est optionnel en GSM. Ce codec voix à débit adaptatif propose huit débits différents allant de 4,75 kbps à 12,2kbps, contrôlés par le réseau d'accès et ne dépendant pas de l'activité vocale. Le codeur AMR fonctionne sur des trames de voix de 20 ms, correspondant à 160 échantillons, échantillonnés à une fréquence de 8khz (théorème de Shannon), le système de codage employé par le codeur AMR est appelée ACELP (algebraic code excited linear prediction coder). Par extension, le codeur ACELP à un débit adaptatif est appelé MR-ACELP (*multi-rate ACELP*). Durant une conversation téléphonique, chaque participant parle en moyenne la moitié du temps, chaque sens de transmission est utilisé donc 50% du temps. La technique AMR prend en compte cette caractéristique spécifique des communications vocales en proposant les quatre fonctions suivantes :

- détection de l'activité vocale au niveau de l'émetteur ;
- évaluation du bruit de fond au niveau de l'émetteur afin de transmettre les paramètres correspondants au destinataire ;
- transmission du bruit de " confort " au destinataire au moyen d'une trame spécifique, envoyée périodiquement ;

- génération du bruit de confort au niveau du destinataire durant les périodes pendant lesquelles aucune trame de voix n'est reçue. La technique AMR possède une fonctionnalité de dissimulation des erreurs (*error concealment*). Le but de substitution de trames est de masquer les effets de la perte de trames. Le débit peut être modifié par le réseau d'accès en fonction de charge de l'interface air et de la qualité de bout en bout des communications vocales, aux heures chargées, pendant lesquelles la charge du réseau est forte. Il est possible d'utiliser un débit AMR plus faible afin d'offrir une capacité plus importante, tout en réduisant quelque peu la qualité vocale. D'autre part, lorsqu'un mobile quitte la zone de couverture d'une cellule et qu'il utilise déjà sa puissance maximale de transmission, un débit moins important peut être utilisé pour étendre artificiellement la zone de couverture de la cellule. Grâce à ce codec AMR, il est possible de réaliser un compromis entre la capacité du réseau, la couverture et la qualité vocale en fonction des besoins des opérateurs.

Classes de services et QoS définies par l'UMTS : Le 3GPP (*3rd Generation Partner Project*) (3GPP, 2000) a défini les classes de services que l'UMTS peut offrir et qui dépendent de la QoS (*Quality of Service*) désirée par l'utilisateur. La qualité de service est généralement définie par les critères suivants :

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

Ces classes de services sont définies par 3GPP comme suit :

1. *services conversationnels (Conversational class)* : les services de cette catégorie permettent de transmettre principalement du son et des images entre deux personnes ou groupes de personnes. Leur qualité est donc assez sensible à des délais dans la transmission, ces délais pouvant être facilement perceptibles par les utilisateurs. Parmi ces services, nous pouvons citer la téléphonie en mode circuit, la téléphonie en mode paquet utilisant le protocole VoIP (Voice over IP) et la vidéophonie. Néanmoins, cette dernière est moins tolérante aux erreurs que la téléphonie et requiert généralement un débit plus important ;
2. *services streaming (Streaming class)* : on retrouve dans cette classe les services qui traitent l'information au niveau du terminal de façon continue, c'est à dire au moment même où elle est reçue (avant de la recevoir dans sa totalité). Ce type de service est unidirectionnel. Ces services sont assez récents et offrent un potentiel d'application assez vaste. Le streaming audio et vidéo sont des exemples d'applications typiques. À la différence des services conversationnels, les services de type streaming sont assez asymétriques (on ne reçoit pas au même débit que l'on émet) et on peut tolérer dans ce cas des délais importants ;

3. *services interactifs (Interactive class)* : ce type de service est caractérisé par le fait que le terminal ayant envoyé un message/commande attend une réponse du destinataire dans un certain intervalle de temps. Par conséquent, un paramètre clé pour identifier la QoS des services interactifs est le temps d'aller-retour, c'est à dire le temps écoulé entre le moment où le message/commande est émis et celui où la réponse arrive. Comme exemple, nous pouvons citer le commerce électronique ainsi que les services de localisation qui se développent de plus en plus ;
4. *services en arrière-plan (Background class)* : il existe des services qui ne posent, ou presque, aucune contrainte de temps dans le transfert de l'information qu'ils génèrent. Ils peuvent s'exécuter en «arrière-plan». Parmi ces services, nous pouvons trouver l'envoi d'un courrier électronique, d'un message court SMS (*Short Message Service*) ou alors le téléchargement d'un fichier. Le paramètre pour quantifier la QoS de ce type de service est principalement le taux d'erreurs qui doit être souvent assez faible.

Architecture de l'UMTS : Le réseau UMTS repose sur une architecture flexible et modulaire. Cette architecture n'est associée ni à une technique d'accès radio, ni à un ensemble prédéfini de services, ce qui assure sa compatibilité avec d'autres réseaux mobiles de deuxième et troisième génération et garantit son évolution. L'architecture peut être adaptée en fonction des besoins des opérateurs de télécommunications. Une telle architecture, comme l'illustre la figure 1.6, est composée de trois «domaines» : le domaine de l'équipement de l'utilisateur UE (*User Equipment*), celui du réseau d'accès radio «universel» UTRAN (*Universal Terrestrial Radio Access Network*) et celui du réseau coeur CN (*Core Network*) (Holma et Toskala, 2001; Javier et Mamadou, 2004). Les domaines du réseau d'accès et du réseau coeur constituent le domaine de l'infrastructure. Chaque domaine réalise une fonction bien précise dans le réseau, tandis que des points d'échange, dénotés par U_u et I_u , servent d'interfaces permettant les échanges entre les différentes parties du réseau.

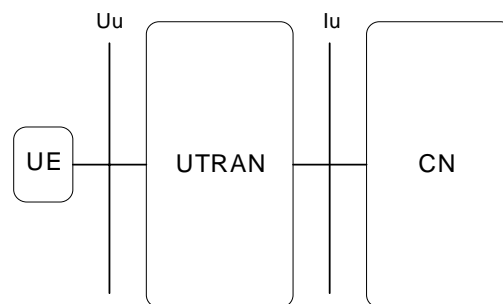


FIGURE 1.6 – Architecture globale du réseau UMTS.

Le domaine UE comprend tous les équipements terminaux et permet à l'abonné l'accès à l'infrastructure et à ses services par le biais de l'interface U_u . Le domaine UE est constitué de deux domaines : celui de l'équipement mobile ME (*Mobile Equipment*) et

celui du modèle d'identité des services de l'abonné ou carte USIM (*Universal Subscriber Identity Module*).

Le domaine UTRAN est composé d'un ensemble de sous-systèmes nommés RNS (*Radio Network Subsystem*) et peuvent être connectés entre eux grâce à l'interface Iur, comme le montre la figure 1.7. Le RNS se scinde en deux éléments : celui du Node-B (ou plusieurs Node-B) et celui du RNC (*Radio Controller Network*) où s'interface avec le domaine CN, généralement à un MSC et à un SGSN grâce aux interfaces Iu CS et Iu PS. Le RNC commande un ou plusieurs Node-B via l'interface Iub. Le Node-B gère principalement la couche physique de l'interface air (le codage du canal, l'entrelacement, l'adaptation du débit et l'étalement). L'UTRAN fournit à l'UE les ressources radio et les mécanismes nécessaires pour accéder au réseau coeur. Il contrôle les fonctions liées à la mobilité et à l'accès au réseau.

Enfin, le réseau coeur CN regroupe l'ensemble des équipements assurant les fonctions telles que le contrôle d'appels, le contrôle de la sécurité et la gestion de l'interface avec les réseaux externes. Ces derniers se scindent en deux catégories : celle des réseaux circuit (Réseau Numérique à Intégration de Service RNIS) et celle des réseaux paquet (réseau Internet), voir la figure 1.7.

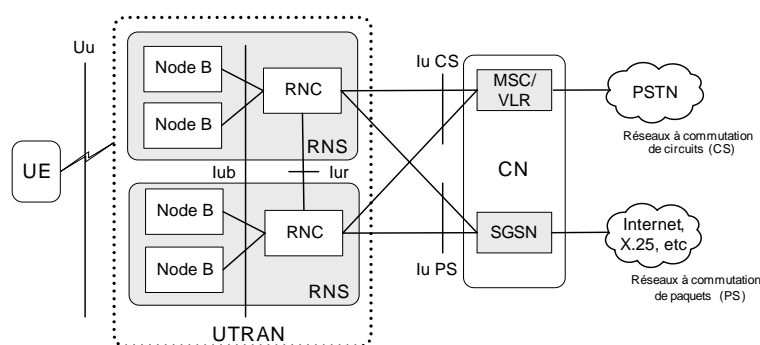


FIGURE 1.7 – Éléments de réseau d'accès UMTS (Holma et Toskala, 2001).

1.1.4 Présentation de HSDPA

Pour offrir des services (de type *interactive, streaming et background*) à haut débit supérieurs à 2 Mbits/s par l'UMTS, le 3GPP a défini HSDPA (*High Speed Downlink Packet Access*) dans la *Release 5*. Il s'agit d'une technologie d'accès radio qui permet d'atteindre des hauts débits dans le lien descendant supérieurs à 10 Mbits/s. La technique HSDPA n'est que l'évolution logicielle de la technologie WCDMA de la *Release 99*. Elle est dotée d'un ensemble de propriétés dont la combinaison permet d'améliorer la capacité du réseau ainsi que le débit de données jusqu'à plus de 10 Mbits/s. Parmi ces propriétés, nous retrouvons des techniques connues et utilisées dans des standards d'évolution tel que le GSM/EDGE et que l'on résume dans la figure 1.8 et l'on définit par les points suivants :

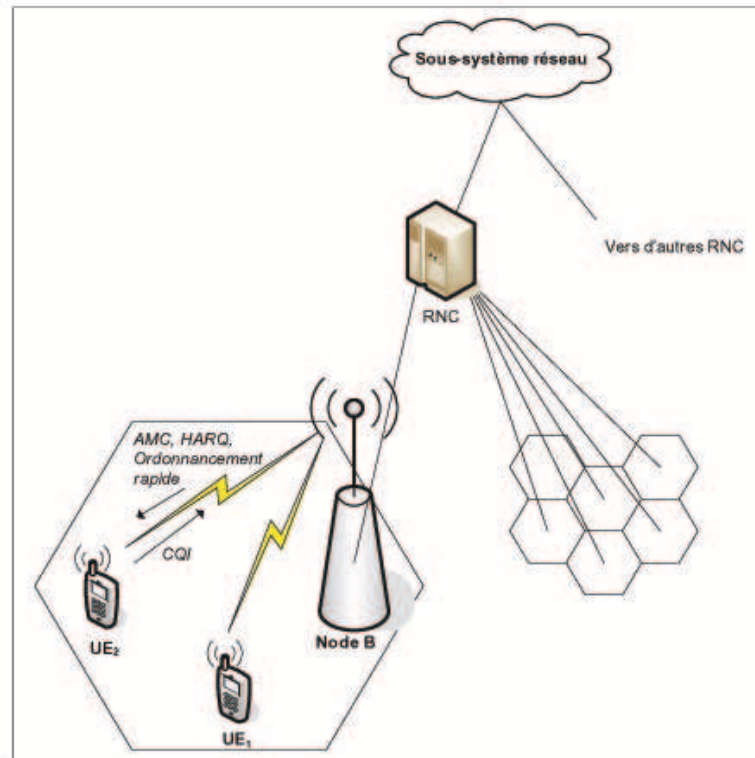


FIGURE 1.8 – Architecture système avec le HSDPA.

1. la technique de modulation et de codage adaptatif AMC permet la variation du débit de transmission de manière adaptative pour pallier les dégradations du signal dues aux conditions de propagation. Cependant, les performances de cette technique sont assez sensibles à des erreurs dans l'estimation des conditions du canal et aux retards implicites de sa transmission vers le Noeud-B;
2. une méthode de retransmission hybride rapide appelée H-ARQ (*Hybrid Automatic Repeat reQuest*) : la méthode H-ARQ est vue comme un complément à la précédente AMC en apportant la possibilité d'ajuster le débit de transmission de manière plus fine. Le Node-B transmet un paquet de données au mobile. Si au bout d'un certain temps ce dernier n'envoie pas un acquittement positif (ACK, *Acknowledgement*) ou si l'acquittement est négatif (NACK, *Negative-Acknowledgement*) alors, le Noeud-B considère que le paquet n'a pas été reçu convenablement et il renvoie à nouveau le même paquet. Le mobile le garde et le combine avec les paquets retransmis par la suite. Ce type de retransmission est appelée *soft combining* et il existe un autre type qui s'appelle *Incremental Redundancy* (Javier et Mamadou, 2004). Ce qui augmente la probabilité de décoder correctement l'information ;
3. des algorithmes d'ordonnancement rapide de paquets (*Fast Packet Scheduling*) : les deux techniques précédentes permettent d'améliorer les performances de la liaison radio en faisant changer le débit de transmission en fonction des caractéristiques instantanées du canal. Les algorithmes d'ordonnancement permettent

de sélectionner les utilisateurs de cellule auxquels il faut allouer le canal de transmission de donnée HS-DSCH (*High Speed-Downlink Shared Channel*) pendant un interval de temps de transmission ITT (*Interval Time Transmission*). Parmi les stratégies d'allocation de ressources radio appelées *scheduling* on trouve le Max C/I, le PF (*Proportional Fair*) et le FFTH (*Fast Fair Throughput*), pour plus de détails sur ces algorithmes voir ([Ameigeiras et al., 2004](#)).

1.1.5 Présentation de IEEE 802.16e (WiMAX)

Le Wimax est une technologie principalement utilisée pour les MAN (*Metropolitan Area Network*). Il permet, notamment, aux zones rurales de se doter d'une connexion internet haut débit. L'objectif du WiMAX est de fournir une connexion internet à haut débit sur une zone de couverture de plusieurs kilomètres de rayon. Les débits théoriques du WiMAX sont de 70 Mbits/s avec une portée de 50 kilomètres. En pratique dans le pire des cas, c'est-à-dire en NLOS «*Non Line Of Sight*» le débit atteint 12 Mbits/s jusqu'à 4,5 kilomètres. Le WiMAX fonctionne en mode point-multipoint, c'est-à-dire le mode infrastructure que l'on connaît pour le Wifi ou encore le même fonctionnement que les technologies 2G, 3G de téléphonie mobile. Ainsi comme en 2G une station de base nommée BTS (Base Transceiver Station) ou BS (Base Station) émet vers les clients et réceptionne leurs requêtes puis les transmet vers le réseau du fournisseur d'accès. Plusieurs variantes de standard ont été proposées, modifiées et ratifiées comme le résume le tableau 1.2.

Standard IEEE802.16	Bande de fréquence (BF)	Date
IEEE std 802.16	réseaux métropolitains sans fil BF supérieure à 10 GHz.	octobre 2002 obsolète
IEEE std 802.16a	réseaux métropolitains sans fil BF comprise entre 2 et 11 GHz.	octobre 2003 obsolète
IEEE 802.16b	réseaux métropolitains sans fil BF comprise entre 10 et 60 GHz.	fusionné avec 802.16a
IEEE std 802.16c	réseaux métropolitains sans fil dans les bandes BF libres.	juillet 2003
IEEE 802.16d (IEEE std 802.16-2004)	révision intégrant les standards 802.16, 802.16a et 802.16c.	octobre 2004
IEEE std 802.16e	réseaux métropolitains sans fil avec des clients mobiles.	février 2006
IEEE std 802.16f	réseaux sans fil maillés (mesh network).	

TABLE 1.2 – Historique du standard IEEE802.16

Signalisation montante de contrôle : Nous nous intéressons par le standard IEEE802.16e ([IEEE Std 802.16e, 2005](#)) qui utilise dans sa couche physique la méthode OFDMA initialement développée par Sari et Karam ([Sari et Karam, 1995](#)) comme une méthode d'ac-

çes multiple. Afin de signaler des événements spécifiques, les mobiles émettent des signalisations (*Ranging*) en direction de la station de base. Ces signalisations sont de différentes natures en fonction du type d'événement à signaler. La transmission des signalisations utilise un ensemble de 256 codes orthogonaux (*codes CDMA*) codés sur 144 bits. Ces codes sont distribués entre les quatre types de signalisation possibles. Un mobile désirent émettre une de ces signalisations choisit aléatoirement un des codes disponibles dans le sous-ensemble propre à sa famille de signalisation. Ce code est alors transmis via une modulation par phase binaire (BPSK). Le standard IEEE802.16e établit quatre catégories de signalisation : l'initialisation de connexion IR (*Initial Ranging*), la demande de migration de cellule HR (*Hand-Over Ranging*), la demande de bande passante BR (*Bandwidth Request Ranging*) et la demande périodique PR (*Periodic Ranging*). Ces signalisations sont définies comme suit :

1. initialisation de connexion IR (*Initial Ranging*) : l'initialisation de connexion est utilisée par un mobile dès lors qu'elle désire se synchroniser pour la première fois à une station de base. Ce message est envoyé sur un premier intervalle de contention : intervalle de signalisation d'initialisation (*Initial Ranging Interval*) ;
2. demande de migration de cellule HR (*Hand-Over Ranging*) : ce type de *Ranging* demande l'engagement d'une procédure de *Hand-Over* afin que le mobile puisse se rattacher à une autre BS tout en conservant la pérennité des services en cours. Ce message est envoyé sur le même intervalle que ceux d'initialisation de connexion ;
3. demande de bande passante BR (*Bandwidth Request Ranging*) : ce type de *Ranging* est la première étape à l'engagement d'un flux de service vers la station de base. Par ce type de requête, le mobile demande à la station de base de lui allouer un nombre déterminé de fréquences et d'unités de temps qui définiront son sous-canal de transmission. La plage temps-fréquence demandée, et les détails de la connexion dépendront du type de service voulu ;
4. demande périodique PR (*Periodic Ranging*) : ce type un peu particulier constitue des signalisations de contrôle. Le mobile est amené à émettre ce type de signalisation sur demande de la station de base afin de pouvoir engager des opérations de maintenance : adaptation du débit, mesure de puissance, changement de modulation, . . .

Classes de services et QoS définies par le standard IEEE802.16 : Les services d'ordonnancement consistent en les mécanismes supportés par la couche MAC (*Medium Access Control layer*) pour le transport de données. Chaque connexion est associée à une unique classe de service. Ces classes de service impliquent un ensemble particulier de paramètres définissant la QoS associée à la connexion. Quatre classes de services sont proposées dans le standard ([IEEE Std 802.16, 2004](#); [IEEE Std 802.16e, 2005](#)) : le service à acceptation non-sollicitée (UGS), le service temps réel (rtPS), le service de transfert de données (nrtPS) et trafic interactif (BE). En outre, une dernière classe a récemment intégré le standard constituant une extension du service temps réel (ertPS). Elles sont définies comme suit :

1. service à acceptation non-sollicitée UGS (*Unsolicited Grant Services*) : la classe UGS

est destinée à supporter les flux de données temps réel caractérisés par une taille fixe de paquets reçus périodiquement. Cette classe est donc destinée aux services T1/E1 et la voix sur IP (VoIP) sans suppression de silence. Les paramètres de QoS associés sont alors le débit maximum et minimum à réserver pour le trafic (souvent les mêmes), le délai maximum, la gigue tolérée, la politique de mise à jour. Le mobile conserve un délai optimal par le fait qu'aucune requête n'est envoyée dans le sous canal de contention une fois le service engagé. Cette classe de trafic élimine donc les données et délais induits par des requêtes de bande passante répétées. La BS réserve alors périodiquement au mobile une partie des ressources radio disponibles. La quantité réservée est initialement égale au débit maximum demandé. Cette classe est dédiée aux flux en temps réel. Les trames sont de taille fixe et émettent régulièrement ;

2. service temps réel rtPS (*real-time Polling Services*) : la classe rtPS est destinée à traiter les trafics temps réel pour lesquels la taille des paquets de flux de données est variable et à intervalle régulier. Aussi, cette classe est utilisée pour les trafics vidéo tels que le MPEG, H.263, . Les paramètres de Qualité de Service s'articulent autour d'un débit minimum réservé pour le trafic vidéo, d'un débit maximum acceptable, d'un délai maximum et enfin d'une périodicité de mise à jour. Le mobile engageant un trafic rtPS a la possibilité de mettre à jour ses paramètres de performances en incorporant périodiquement de nouvelles requêtes de ressources en queue des paquets de données. Durant l'intervalle de contention. Cette classe respecte donc les trafics sensibles au délai en évitant aux requêtes de mise à jour, les collisions impliquées par l'intervalle de contention. Par ailleurs, le mobile a tout loisir d'indiquer à la BS quelle périodicité de mise à jour requière le trafic engagé. Elle permet les flux temps réel. Les trames sont de taille variable et l'émission peut être irrégulière. Cette classe convient, par exemple, à un flux vidéo ;
3. extension du service temps réel ertPS (*extended real time Polling Service*) : la classe ertPS est destinée à supporter les flux de données temps réel caractérisés par une taille variable de paquets reçus périodiquement. Cette classe est donc destinée aux services de voix sur IP (VoIP) avec suppression de silence. Les paramètres de Qualité de Service s'articulent autour d'un débit minimum réservé pour le trafic vidéo, d'un débit maximum acceptable, d'un délai maximum mais sans indicateur de périodicité de mise à jour. La SS engageant un trafic ertPS a la possibilité de mettre à jour ses paramètres de performances en incorporant, quand cela est nécessaire, de nouvelles requêtes de ressources en queue des paquets de données ou durant l'intervalle de contention. Cette classe respecte donc les trafics sensibles au délai en évitant aux requêtes de mise à jour les collisions impliquées par l'intervalle de contention. Par ailleurs, cette classe de trafic réduit le trafic de signalisation et le délai d'évolution pour les ressources allouées dynamiquement. De plus, l'allocation variable des ressources pour les appels ertPS permet d'éviter le gaspillage de ressource sur lien montant, comme cela peut être le cas sur les appels UGS ;
4. service de transfert de données nrtPS (*non real time Polling Service*) : la classe nrtPS doit quant à elle permettre la bonne gestion des trafics insensibles au délai mais requérant un débit minimum. La taille des paquets peut donc être variable, ainsi

que le délai entre deux transmissions de paquets. Cette classe est donc tout à fait adéquate au protocole de transfert de fichiers (FTP). Les paramètres de Qualité de Service sont : un débit minimum de trafic à réserver, le débit maximum possible pour la connexion et une priorité de trafic. Une fois la connexion engagée, la SS devra impérativement émettre une nouvelle requête de bande passante pour pouvoir prétendre à un autre débit ;

5. service interactif BE (*Best Effort*) : la classe BE est dédiée aux trafic n'exigeant pas de niveau de performance particulier. Les paramètres de QoS sont uniquement liés au maximum de débit possible pour la connexion, à la priorité de trafic et à la politique de renouvellement de requête. Comme pour la classe précédente, la classe engage une nouvelle connexion pour tout nouveau besoin de ressource. Dans cette classe rien n'est garanti on envoie les données sans garantie de réception. Convient bien à la navigation Internet.

1.2 Travaux existants

Cette partie décrit les travaux de littérature déjà existants dans lequel le travail engagé dans cette thèse tire ses motivations. Elle traite des travaux proposant différents modèles de performances, de calcul de capacité, de gestion de mobilité et de mécanisme de contrôle d'admission dans les axes de recherche étudiés le long de cette thèse qui concernent les systèmes WCDMA, HSDPA et WiMAX. Certains de ces travaux présentent des formulations théoriques adéquates et éprouvées aux enjeux de notre présente étude. D'autres se caractérisent par leurs conclusions qui contribuent fortement à l'approfondissement des nouvelles pistes de recherche ayant motivé cette thèse. Ainsi d'autres travaux associés aux modèles de capacité et de contrôle d'admission abordent les principaux moyens et outils de modélisation de la capacité d'une cellule HSDPA et IEEE802.16 avec ou sans la gestion des ressources par rapport à la mobilité des utilisateurs.

De nombreux travaux ont été développés dans la littérature pour étudier la capacité des réseaux sans fil. Dans (Gilhousen et al., 1990), Gilhousen et al. analysent les interférences venant des autres cellules sur la capacité multi-cellulaires CDMA dans laquelle il y a un même nombre constant de mobiles par cellule qui se distribue de façon continue dans chacune d'elles. Dans (Viterbi et Viterbi, 1993), les auteurs étendent le travail développé dans (Gilhousen et al., 1990) en modélisant le système CDMA par une file d'attente de type $M/G/\infty$ dont le nombre de mobiles actifs par cellule est un variable aléatoire qui suit la loi de Poisson dont le paramètre est le trafic moyen offert dans chaque cellule. Cependant, les auteurs utilisent les limites de Chernoff et les approximations gaussiennes pour le calcul de la probabilité de mauvaise liaison (*outage probability*). Également, ils prennent en considération l'effet des interférences inter-cellulaires sur les performances du système CDMA.

Dans (Evans et Everitt, 1999), Evans et al. analysent la capacité du trafic du lien montant des réseaux cellulaires CDMA en incluant les interférences venant des mobiles d'autres

cellules sous forme de puissances comme des points intéressants dans leur développement afin qu'ils analysent le trafic. Chaque mobile a une position aléatoire à l'intérieur de sa cellule qui génère une interférence comme variable aléatoire sous forme de puissance sur les autres mobiles au niveau de leurs station de base, le nombre de mobile par cellule est un variable aléatoire suit la loi de Poisson au lieu d'être un nombre constant. Ensuite, ils modélisent le système CDMA comme une file d'attente $M/G/\infty$ et donnent une nouvelle limite de la capacité CDMA en terme de probabilité de mauvaise liaison. Ils calculent cette probabilité par les deux techniques d'approximation : celle de l'approximation gaussiennes et celle du calcul des limites de Chernoff. Ils considèrent dans leur analyse l'effet des interférences de type *shadowing lognormal* et le contrôle de puissance. Signalons que les travaux mentionnés précédemment ne traitent que le service voix. Donc, offrir d'autres services comme le multimédia par le réseau CDMA, demande une qualité de service acceptable par les mobiles.

Les auteurs dans (Zhang et Yue, 2001) présentent une méthode pour calculer la capacité d'Erlang du lien montant (*uplink*) du système WCDMA en se basant sur le modèle LCH (*Lost Call Held*) décrit dans (Viterbi, 1995). Les auteurs prennent en considération différents facteurs, tels que le niveau d'interférence, les erreurs de contrôle de puissance, le rapport signal sur interférences plus bruit, etc. Cette méthode calcule l'occupation de la capacité du système UMTS/WCDMA sur la base de la probabilité de *outage*. Ainsi, les auteurs donnent une forme explicite de la capacité d'Erlang pour un seul type de trafic (voix) et également pour des trafics mixtes (voix et données). En outre, dans les travaux développés dans (Mandayam et al., 1997), la capacité du lien montant est étudiée en présence de deux trafics : le trafic en temps réel est transmis continuellement, et le non-temps réel est transmis en temps partagé. Dans (Altman, 2002), l'auteur calcule la capacité du lien montant de systèmes CDMA avec un contrôle de puissance idéalisé qui contient des applications de type *best-effort*, c'est à dire des applications dont le taux de transmission peut être contrôlé. Un appel de type *best effort* qui arrive est supposé avoir une quantité fixe de données à transmettre, ce qui fait que le débit de transmission qui lui est alloué par le système détermine la durée de cet appel. L'auteur permet une variété de services et donc des terminaux mobiles peuvent avoir des qualités requises de services différentes selon le type d'appel. Il garantit la qualité de service d'appels en cours en supposant qu'il y a un contrôle d'accès exercé afin d'empêcher un nouvel appel d'arriver quand le système est saturé. Les auteurs dans (Hegde et Altman, 2003) étendent la notion de capacité définie dans (Altman, 2002) à d'autres type de la QoS. Ils considèrent deux types de trafics à savoir, le trafic temps réel et *best effort*. Puis, ils étudient l'impact de la bande passante fixée par appel de type *best effort* sur la capacité d'Erlang du système et montrent qu'une très faible bande passante réservée aux appels *best effort* indique une durée d'appels très longue. La capacité correspondante aux appels *best effort* pour un certain temps de séjour est définie comme le taux d'arrivée de ces appels que le système peut gérer avec la contrainte que leurs délais moyens de transmission prévue ne dépassent pas un seuil fixé par l'opérateur de télécoms.

Dans (Kelif et Altman, 2004), les auteurs considèrent le système WCDMA en présence de deux trafics, temps réel (RT) et non temps réel (NRT) . Ils supposent que les appels en RT ont des ressources dédiées par le système, par contre les appels en NRT partagent

des ressources libres. Ils étudient la probabilité de blocage des nouveaux appels du trafic RT et le temps de séjour prévu pour les appels du trafic NRT sur les deux liens de transmission montant et descendant (*uplink et downlink*) avec et sans macro diversité. Puis, ils proposent un contrôle d'admission CAC (*call admission control*) afin d'avoir une QoS pour les deux trafics et donnent une extension de leurs travaux en incluant le *handover*.

Dans (Bonald et Proutière, 2005) un modèle général a été proposé pour analyser la capacité des réseaux cellulaires en se basant sur une prédiction par la théorie de l'information. Ce travail est premièrement applicable dans les systèmes qui utilisent la méthode de multiplexage temporelle (*time-multiplexed*) comme CDMA 1xEV-DO, UMTS HSDPA et fréquentielle (*frequency-multiplexed*) comme OFDM. Une comparaison a été effectuée avec la capacité maximale obtenue par la théorie de l'information. Le modèle décrit une relation entre les taux de transmission de données et la quantité des ressources utilisées en terme du niveau de puissance dans les systèmes cités ci-dessus. Les auteurs présentent des résultats qui accentuent l'importance de modéliser le système en prenant en considération les arrivées et les départs dynamiques d'appels. La principale contribution de ces auteurs est le développement d'une relation récursive entre les états du système. En se basant sur cette relation pour obtenir les limites inférieures et supérieures, des probabilités de blocage sont établies quand le contrôle d'admission est basé sur un taux minimum de transmission de données.

Dans (Fodor et Telek, 2005), les auteurs définissent une nouvelle méthode en étendant la formule classique de Kaufman-Roberts dans le système CDMA supportant des services *best effort*. Les services de cette classe peuvent adapter dynamiquement leur taux de transmission en fonction des interférences. Les auteurs utilisent la formule de Kaufman-Roberts pour calculer l'état d'équilibre en fixant des taux de transmissions dans le système. Ils donnent une approximation d'une chaîne de Markov irréversible par celle réversible et obtiennent des limites inférieures et supérieures de probabilités de blocages des nouveaux appels dans l'état d'équilibre.

Le CAC décide d'accepter un nouvel appel s'il ne dégrade pas la QoS des appels en cours (Tang et Goldsmith, 2001; Leong et Zhuang, 2002). Le CAC dans le réseau de troisième génération UMTS a fait également l'objet de plusieurs travaux de recherche (Elayoubi et al., 2004, 2005, 2006) pour le lien montant et descendant. Les auteurs ont en particulier calculé la capacité du système pour trois types de récepteurs, à savoir le Filtre Adapté (*Matched Filter*), MMSE (*Minimum Mean Square Error*) et le Décorrélateur (*Decorrelator*). En plus du rapport signal sur interférence plus bruit SINR (*signal to noise plus interference ratio*), ils prennent en compte d'autres contraintes, telles que la couverture et la mobilité. Ils ont également mis en oeuvre plusieurs algorithmes de contrôle d'admission d'appels, qui gèrent les priorités entre les nouveaux appels et ceux en *handover* ou en mobilité, tout en considérant le trafic voix et données dans leur étude. Ils ont utilisé deux approches : celle proactive, fondée sur mesure et celle réactive, basée sur la compression du trafic de données. Tous ces travaux nous ont encouragé à proposer d'autres méthodes efficaces qui nous permettent de trouver explicitement les performances du système (voir les chapitres 2 et 3).

Dans (Assaad, 2006), Assaad et al étudient l'impact du TCP (Transmission Control Protocol) sur la capacité et la performance du système UMTS-HSDPA. Les auteurs proposent une méthode analytique pour l'évaluation de cet impact et une méthode de réduction de l'effet du protocole TCP relatif au débit binaire sur le réseau mobile sans fil UMTS en utilisant des canaux partagés.

Puisque la qualité du signal varie en fonction de la distance du mobile par rapport à sa station de base, et donc le mobile aura un débit qui se dégrade quand il s'éloigne de la station de base. Plusieurs chercheurs ont axé leur étude sur ce problème en subdivisant la zone couverte par la cellule en un nombre fini de région concentriques. Dans (Bonald et Proutière, 2003), les auteurs considèrent un réseau sans fil de données où le nombre de mobiles actifs est dynamique et les mobiles partagent les ressources radio du lien descendant dans les systèmes CDMA/HDR. Ils développent analytiquement les performances du système comme la probabilité de blocage et le débit total, ainsi que la taille de la cellule et la densité du trafic, tout en se basant sur le dimensionnement de la zone couverte par la station de base en un nombre fini de régions.

Dans (Chahed et al., 2008), les auteurs présentent un modèle pour étudier la capacité du système HSPA en combinant les deux liens de transmissions montant et descendant avec la présence de deux classes de service temps réel et *best effort*. La dernière classe est gouvernée par le protocole TCP dans la couche de transport et dont les paquets sont transmis sur le lien descendant et ACKs sur le lien montant. Les appels temps réels ont des ressources dédiées et ceux en *best effort* partagent équitablement les ressources libres. Les appels *best effort* restent longtemps dans le système s'il y a moins de ressources libres et le quittent rapidement quand ils ont plus de ressources libres. Les auteurs analysent le système HDR/HSDPA en offrant les performances du système et proposent un contrôle d'admission pour les appels *best effort* dans les deux liens de transmission.

Les auteurs dans (Li et Sampalli, 2007) développent un contrôle d'admission sur lien descendant du réseau haut débit WiMAX et HSDPA en se basant sur l'adaptation du lien radio de l'interface air. Ils décomposent la cellule en nombre fini de régions prenant en compte un modèle simple de propagation opérant avec uniquement un *path loss* en fonction de la distance entre le mobile et la station de base. Ils considèrent dans leur étude une seule classe de service et modélisent le système par le réseau de files d'attente de type BCMP (Baskett et al., 1975; Chao et al., 2001) dont le nombre d'utilisateurs actifs est fixe. Les auteurs supposent que les taux d'arrivée d'appels dans chaque région uniquement en fonction du nombre moyen de mobile qui se déplacent dans cette région. Ainsi, ils considèrent que le temps moyen de service dépend seulement du temps que le mobile reste dans la région.

Les auteurs dans (Horrich et al.) s'intéressent à la gestion de ressource des réseaux HSDPA et WiMAX en tenant compte de la mobilité. Ils présentent trois schémas RRM (Radio Resource Management) de gestion de ressources dans un système coopératif contenant HSDPA et WiMAX et comparent les performances dans ces schémas en termes de probabilité de blocage des nouveaux appels, de perte d'appels en cours et de temps de transfert de fichiers. Les auteurs se basent sur un modèle de Markov avec la

théorie des jeux pour la gestion de ressources en présence de plusieurs régions d'efficacité de modulation. Ces travaux font la motivation principale de notre contribution développée dans le chapitre 4.

La gestion de la qualité de service et l'évaluation de performances dans la couche MAC du réseau IEEE 802.16e ont été largement effectuées en simulation ; et peu de résultats analytiques ont été prouvés dans la littérature. Dans (Won et al., 2003), les auteurs analysent les performances du protocole d'accès aléatoires dans un environnement OFDMA-CDMA utilisant les intervalles de contention pour l'engagement des connexions, en fonction du délai moyen MDT. Dans (Seo et al., 2004), les auteurs définissent un modèle de performance basé sur un contrôle adaptif de la taille de chaque famille de codes : l'initilisation de connexions (IR), la demande périodique (PR) et la demande de bande passante (BR), afin d'améliorer l'efficacité des accès à compétition. De nombreuses études antérieures partent du principe que les paquets, ou les appels, arrivent dans le système en suivant un processus de Poisson. Ceci, en vue de la simplification du modèle analytique. De plus, ils considèrent que les processus des arrivées de tout type d'appels ou paquets sont indépendents. Néanmoins, puisque dans le WiMAX, tous les mobiles partagent un nombre fini de codes CDMA, les arrivées de ces appels devraient être dépendents, comme prouvé par (Peyre et al., 2008). En effet, dans (Peyre et al., 2008), les auteurs développent un modèle d'accès en couche MAC des requêtes de signalisation en se basant sur les différentes classes de priorités utilisant la différenciation des paramètres *backoff* et les codes libres pour ces classes. Ils considèrent le cas de la classification des connexions basée sur les codes de partitionnement entre les classes d'appels avec une sous plage de codes réservée pour la classe temps réel et une autre sous plage de codes partagée avec celle non temps réel.

Le travail développé dans (Wongthavarawat et Ganz, 2003), définit une architecture de QoS associant un planificateur et un algorithme de contrôle d'admission. De manière plus riche et subtil, l'étude (Wang et al., 2005) développe un algorithme de contrôle d'admission aboutit et dynamique afin d'assurer la qualité de service de chaque classe de trafic du standard IEEE802.16e. L'algorithme réserve sur les ressources disponibles celles demandées par les trafics non temps-réel. Ces derniers seront alors refusés si la capacité de la cellule est dépassée par la nouvelle demande de ressources. En parallèle, les autres types de trafic partageront les ressources restantes. En outre, le principe d'allocation proposé ici introduit une évolution dynamique dite de dégradation de ressource. En effet, l'arrivée de nouveaux appels temps réels aboutira à la réduction progressive et adaptative des ressources restantes partagées entre les appels non temps réel. Ainsi, l'auteur assure la plus grande priorité aux trafics UGS et rtPS tout en maximisant le débit global par une attribution adaptative de la bande passante aux autres trafics. Ce principe, faisant autorité dans la littérature, sert de base à l'élaboration du contrôle d'admission développé au chapitre 5. En outre, nous voyons au travers de (Benameur et al., 2001), une base théorique menant à un partage efficace et équitable des ressources radio disponibles. En effet, ce travail constitue la base des algorithmes évolués de contrôle d'admission. Il propose une gestion indépendante des trafics temps réel et non temps réel. Les premiers reçoivent en priorité les ressources demandés, les autres se partagent équitablement la bande passante laissée libre. Par contre aucun des

deux trafics n'obtiendra une nouvelle connexion si les trafics non temps réel ne disposent plus d'un minimum de bande passante. Ce principe vise à empêcher la famine de ressource pour les trafics non temps réel. Dans une toute autre approche, nous présentons les travaux de (Ball et al., 2005). Ce travail propose un ordonnancement innovant de l'allocation de ressources entre les trafics. Le principe de TRS (ordonnancement par omission temporisée) attribut les ressources en fonction de l'efficacité de modulation de chaque utilisateur. Partant du constat que les utilisateurs ayant le plus mauvais rapport signal bruit (SNR) engendrent un plus grand gaspillage des ressources radio, il se verront allouer peu de ressources de communication et uniquement à des intervalles espacés. Ce mécanisme part du principe que la mobilité des utilisateurs leurs permettront à terme de profiter pleinement des ressources qui leurs sont nécessaires. Ce mécanisme original d'allocation de ressource aboutit à un gain important d'efficacité spectrale. Par conséquent, les communications issues des régions proches sont rapidement servies, et dans le contexte d'un réseau non-saturé, ceci aboutit indirectement à une meilleure prise en charge des communications des régions plus périphériques. Ce travail a la particularité de prendre en compte l'efficacité spectrale des utilisateurs vis à vis des ressources radio qui leurs sont allouées. Ce dernier critère nous semble un des critères primordiaux caractérisant un algorithme d'admission et d'allocation performant.

Deuxième partie

Évaluation des performances dans le réseau UMTS/W-CDMA

Chapitre 2

Évaluation des performances par l'Analyse Spectrale

Sommaire

2.1	Introduction	43
2.2	Cas de deux classes de service	45
2.2.1	Formulation du problème	45
2.2.2	Cas d'une cellule WCDMA isolée	47
2.2.3	Analyse du système	49
2.2.4	Approche par l'analyse spectrale	51
2.2.5	Distribution stationnaire	56
2.2.6	Métriques de performances	58
2.2.7	Cas multi cellules	59
2.2.8	Résultats numériques	60
2.3	Cas de trois classes de service	65
2.3.1	Contrôle d'admission du trafic temps réel	65
2.3.2	Contrôle d'admission du trafic non temps réel	67
2.3.3	Trafic <i>best effort</i> sans contrôle d'admission	69
2.3.4	Évaluation des performances par l'analyse spectrale	71
2.3.5	Illustrations numériques	74
2.4	Conclusion	76

2.1 Introduction

Le système de communications mobile UMTS de troisième génération s'appuie sur la technique d'accès multiple WCDMA de l'interface radio. L'avantage de la troisième génération de réseaux mobiles réside dans le fait qu'elle offre aux utilisateurs une variété de services et permet d'obtenir des hauts débits allant jusqu'à 2Mbps. Parmi ces

services, citons la voix, l'accès Internet à haute vitesse, le téléchargement de fichiers audio et vidéo, la réception de programmes télévisés et la visiophonie. Ces services sont étudiés dans la littérature en deux grandes classes à savoir les appels temps réel et les appels *best effort*. Chaque service a une exigence particulière en terme de qualité de service QoS, ce qui peut affecter la capacité du système WCDMA.

Le travail développé dans ce chapitre s'ajoute aux travaux antérieurs sur la capacité du lien montant (la communication du mobile vers la station de base) du système WCDMA. Pour ce faire, nous avons séparé la diversité de services offerts par ce système, en deux puis en trois classes de trafic. Dans la section 2.2, nous considérons le lien montant du système WCDMA avec deux classes de services, temps réel qui ont des ressources dédiées et *best effort* sans aucune exigence sur la QoS. Notre analyse est basée sur la modélisation du système par une chaîne de Markov à deux dimensions, où la première composante correspond au nombre d'appels temps réel et la deuxième composante correspond au nombre d'appels *best effort*. Pour obtenir la distribution stationnaire du nombre d'appels temps réel et *best effort*, nous utilisons l'approche de l'Analyse spectrale avec un processus quasi naissance et mort QBD (Queija, 2000; Elazouzi et Altman, 2004). En particulier, nous obtenons les probabilités d'états d'équilibre en fonction des valeurs propres et vecteurs propres de certaines matrices finies. Nous obtenons une solution explicite contrairement à celle proposée dans (Hegde et Altman, 2003). Ce résultat nous permettra d'évaluer les performances du système en fonction de valeurs propres et vecteurs propres de certaines matrices. Désormais, les appels temps réel ont besoin d'un taux de transmission minimum à garantir, donc le système ne peut manipuler qu'un nombre fini d'appels temps réel. Toutefois, nous montrons que ces valeurs propres et vecteurs propres sont obtenus indépendamment du nombre d'appels *best effort* qui réduit la complexité de calcul de la distribution stationnaire. Une autre application importante de notre approche par analyse spectrale (Pla et Casares-Giner, 2005; Mitrani, 1995; Elazouzi et Altman, 2004; Queija, 2000; Chakka, 1998; Chakka et Mitrani, 1995b,a) est de réduire la complexité de calcul du contrôle d'admission pour les appels *best effort*. Par conséquent, notre approche est utile pour un opérateur dans sa politique d'admission d'appels *best effort*, avec moins de calcul, qui maximise l'utilisation de la bande passante et à garantir la QoS exigée par les appels temps réel et *best effort*.

Dans la section 2.3, nous considérons tous les services supportés par le lien montant du système WCDMA en trois classes à savoir : temps réel, non temps réel et *best effort*. La classe temps réel nécessite un débit constant avec des contraintes strictes sur le délai, comme la voix et la vidéo. La classe non temps réel exige un taux minimum de transmission et peut tolérer un certain délai d'attente, par exemple, le transfert de fichiers FTP. La classe *best effort* quant à elle, considère des applications qui peuvent adapter leur taux de transmission selon les ressources disponibles du réseau, donc les appels de cette classe ne font pas l'objet d'un CAC dans ce travail. La principale qualité de service est de garantir le temps de séjour d'un appel dans le système. Nous introduisons un nouveau mécanisme du CAC pour les appels non temps réel en nous basant sur le temps de séjour qui doit être borné. Le nouvel appel non temps réel est accepté si le délai correspondant au nombre actuel d'appels non temps réel est borné par une constante positive fixée par l'opérateur. Nous supposons que la classe non temps réel

et *best effort* partagent une partie des ressources du réseau avec la classe temps réel. Notre objectif est d'étudier la qualité de service d'appels non temps réel et *best effort* en agissant sur les métriques de performance : le temps de séjour et la probabilité de blocage. En effet, nous proposons un contrôle d'admission qui donne un compromis dans la gestion de la capacité entre les trois classes en terme de qualité de service exigée par chaque appel. La qualité de service est déterminée par rapport au temps de séjour qui doit être borné. L'opérateur peut calculer le taux maximum des arrivées pour assurer ce temps de séjour.

2.2 Cas de deux classes de service

Nous nous intéressons à l'analyse de la capacité WCDMA en nous basant sur l'approche de l'analyse spectrale en présence de deux types de trafic :

- Trafic *best effort* (BE) où aucune garantie n'est faite sur le service offert, autre que l'engagement du système à faire "du mieux qu'il peut" ;
- Trafic temps réel (RT) avec des contraintes de type temps réel et une garantie de délai et de débit.

2.2.1 Formulation du problème

La diversité de service qu'offre le réseau UMTS, aux utilisateurs mobiles, est un axe de recherche qui intéresse plusieurs chercheurs. L'interface air WCDMA a été conçue de telle façon que certaines fonctionnalités avancées du CDMA, comme la détection multi-utilisateurs et les antennes adaptatives, peuvent être déployées par l'opérateur afin d'accroître la capacité et/ou la couverture de son réseau. Cependant, la gestion dynamique de la capacité des systèmes cellulaires CDMA reste un défi majeur à relever par les opérateurs Télécoms. Notons que le nombre d'appels actifs ne donne pas la capacité réelle du système, car elle est limitée par les interférences. Ces interférences peuvent être due au bruit thermique, interférences provoquées par les utilisateurs, qu'elles soient situées dans la même cellule (interférence intracellulaire) ou dans les cellules voisines (interférence intercellulaire), pertes de propagation dues à la distance (*path-loss*), atténuations induites par les obstacles (*shadowing*) et évanouissements (*fading*) suscités par l'existence de trajets multiples. Pour formaliser le problème, nous considérons C classes de services. Cependant, pour un mobile de la classe i ($i \in S = \{1, \dots, C\}$), son rapport signal sur bruit plus les interférences (SIR) à la station de base, doit satisfaire la condition ci-après, afin de lui garantir une communication sans interruption (Laiho et Wacker, 2001) :

$$\frac{P_i}{N + I_{own} + I_{other} - P_i} = \alpha_i \geq \left(\frac{E_i R_i}{N_0 W} \right) = \chi, \quad (2.1)$$

où P_i est la puissance reçue d'un mobile de la classe i par la station de base ; N est la densité de puissance du bruit de fond ; E_i est l'énergie par bit transmis de type i , R_i est le taux de transmission de la classe de service i ; W est l'étalement du spectre de

la bande passante; I_{own} désigne l'intra-interférence reçue de tous les mobiles appartenant à la même cellule et I_{other} désigne l'inter-interférence reçue de tous les mobiles des autres secteurs. Nous rappelons que les interférences inter et intra cellulaires sont respectivement :

$$I_{own} = \sum_{j=1}^C M_j P_j, \quad (2.2)$$

$$I_{other} = g I_{own}, \quad (2.3)$$

où g est la constante des interférences donnée par expérience (Laiho et Wacker, 2001) et M_i est le nombre de mobiles de la classe i . Dans le système WCDMA il existe d'autres sources de dégradation de la qualité du signal tel que *shadowing fading* qui est due aux chemins multiples et aux effets de masque. Donc, le seuil n'est pas toujours fixe c'est-à-dire n'est pas toujours égale à $(\frac{E_i R_i}{N_0 W})$. La variation du *shadow fading* est une variable aléatoire qui suit la distribution lognormal tel que $\alpha_i = 10^{0.1\zeta_i}$, où $\zeta_i \sim N(\mu_\zeta, \sigma_\zeta)$. Afin de prendre en compte ce type de dégradation, les auteurs dans (Hegde et Altman, 2003) ont introduit une nouvelle fonction (Γ), indépendamment du type de classe. Elle dépend uniquement de la valeur de déviation du rapport signal sur interférence α_i telle que

$$\Gamma = 10^{\frac{\sigma_\zeta^2}{20h} - \frac{\sigma_\zeta^2 Q^{-1}(\beta)}{10}},$$

$$\beta = P(\alpha_i > \chi) = \int_{\chi_i}^{\infty} f_{\alpha_i}(x) dx = Q\left(\frac{h \ln \chi - \mu_\zeta}{\sigma_\zeta}\right),$$

$$f_{\alpha_i}(x) = \frac{h}{x \sigma_\zeta \sqrt{2\pi}} \exp\left(-\frac{h(\ln x - \mu_\zeta)^2}{2\sigma_\zeta^2}\right), \quad h = \frac{10}{\ln 10},$$

$$Q(x) = \int_x^{\infty} \frac{e^{-t^2/2}}{2\pi} dt.$$

Donc, le rapport SIR doit être supérieur ou égale à $(\frac{E_i R_i}{N_0 W} \Gamma)$. Pour une meilleure satisfaction d'appels de la classe i , sans dégradation de la qualité de service, et dans un but de servir un nombre important d'utilisateurs, la puissance minimum reçue (P_i) doit satisfaire l'équation suivante

$$P_i = \frac{N\Delta_i}{1 - (1 + g) \sum_{j=1}^C M_j \Delta_j}, \quad (2.4)$$

où $\Delta_i = \frac{\gamma_i}{1 + \gamma_i}$ et $\gamma_i = (\frac{E_i R_i}{N_0 W} \Gamma)$. Le taux de charge du système est défini comme suit

$$\theta = \sum_{j=1}^C M_j \Delta_j. \quad (2.5)$$

Nous obtenons une nouvelle expression de la puissance minimale reçue P_i en fonction du taux de charge

$$P_i = \frac{N\Delta_i}{1 - (1 + g)\theta}. \quad (2.6)$$

Cette puissance doit être finie et positive dans un but de limiter les interférences. D'où le dénominateur qui doit être strictement positive. Notons par Θ_ϵ la capacité du système définie comme étant la borne supérieur du taux de charge θ . Nous avons $\theta \leq \Theta_\epsilon / (1 + g)$ où $\Theta_\epsilon = 1 - \epsilon$ où ϵ est un nombre réel strictement positif. Γ est une fonction de l'écart-type, σ_ζ , du *shadow fading*, dont la valeur varie avec la mobilité des utilisateurs. Dans cette première partie, nous ne considérons pas la mobilité des utilisateurs.

2.2.2 Cas d'une cellule WCDMA isolée

Dans la suite de ce travail, nous limitons notre étude à deux classes de service : temps réel (RT) et *best effort* (BE). Nous considérons une seule cellule WCDMA dans cette section et le cas multi cellulaire sera traité dans la section 2.2.7. Donc, nous négligeons les interférences provenant d'autres cellules, c'est-à-dire, $g = 0$ dans l'équation (2.3).

Nous utilisons le codec AMR¹ pour les appels RT. Le taux minimum de transmission d'appels RT est noté par R_{RT}^m et le taux maximum de transmission est désignée par R_{RT}^r . On définit la bande passante Δ_{RT}^h correspondante au taux R_{RT}^h comme suit

$$\Delta_{RT}^h = \frac{\Gamma E_{RT} / N_0}{W / R_{RT}^h + \Gamma E_{RT} / N_0}, \quad h = r, m, \quad (2.7)$$

où E_{RT} est l'énergie par bit transmis pour RT ; N_0 est la densité du bruit de fond. La capacité disponible pour les BE dépend du nombre d'appels RT. Nous supposons que les appels BE n'ont aucune QoS garantie par le système et donc un nombre infini ou quelconque sera accepté par le système.

Nous modélisons le système WCDMA par une chaîne de Markov à deux dimensions $(X(t), Y(t))$, où $X(t)$ (respectivement $Y(t)$) représente le nombre d'appels de type RT (respectivement BE) dans le système à l'instant t . Les appels de la classe RT (respectivement BE) arrivent selon un processus de Poisson avec le paramètre λ_{RT} (respectivement λ_{BE}) et leurs durées de service distribuées exponentiellement avec le paramètre μ_{RT} (respectivement μ_{BE}). Soit M_{RT} le nombre maximum d'appels RT qui peuvent être servis simultanément par le système, il est donné par $M_{RT} = \lfloor \frac{L_{RT}}{\Delta_{RT}^m} \rfloor$, où $\lfloor x \rfloor$ est la partie entière de x et L_{RT} indique la capacité maximale d'appels RT :

$$L_{RT} = \Theta_\epsilon - L_{BE}, \quad (2.8)$$

où L_{BE} est la proportion minimale de la capacité réservée aux appels BE. Les appels BE peuvent utiliser la bande passante non utilisée par les appels RT comme suit

$$C(X_{RT}) = \begin{cases} \Theta_\epsilon - X_{RT} \Delta_{RT}^r, & \text{si } 1 \leq X_{RT} \leq N_{RT}, \\ L_{BE}, & \text{si } N_{RT} < X_{RT} \leq M_{RT}, \end{cases} \quad (2.9)$$

1. Pour la définition du codeur AMR, voir la référence (Holma et Toskala, 2001)

où N_{RT} est le nombre d'appels RT correspondant à la capacité normalisée (Δ_{RT}^r) requise par chaque appel RT, où $N_{RT} = \left\lfloor \frac{L_{RT}}{\Delta_{RT}^r} \right\rfloor$.

Nous signalons que les appels RT sont indépendants d'appels BE, d'où nous modélisons le système avec seulement les appels RT par une file d'attente $M/M/M_{RT}/M_{RT}$ avec un taux d'arrivée λ_{RT} et un taux de service exponentielle noté μ_{RT} . A partir des équations de balances d'une chaîne de Markov associée à cette file d'attente, les probabilités stationnaires $q_i = \lim_{t \rightarrow \infty} Pr[X(t) = i]$ peuvent être réécrites de manière récursive comme :

$$q_{i+1} = \frac{\rho_{RT}}{i+1} q_i, \quad i = 0, 1, \dots, M_{RT}, \quad (2.10)$$

où $\rho_{RT} = \frac{\lambda_{RT}}{\mu_{RT}}$. Utilisant l'équation (2.10) et la condition de normalisation $\sum_{i=0}^{M_{RT}} q_i = 1$, chaque probabilité d'obtenir i appels RT dans le système est calculée par

$$q_i = \frac{(\rho_{RT})^i / i!}{\sum_{j=0}^{M_{RT}} \frac{(\rho_{RT})^j}{j!}}, \quad (2.11)$$

Alors, la probabilité de blocage d'un nouvel appel RT s'écrit

$$P_B^{RT} = q_{M_{RT}} = \frac{(\rho_{RT})^{M_{RT}} / M_{RT}!}{\sum_{j=0}^{M_{RT}} \frac{(\rho_{RT})^j}{j!}}. \quad (2.12)$$

Les appels BE dépendent de ceux RT en cours. Nous considérons les appels des deux classes simultanément dans notre analyse. Soient $K(t) = (X(t), Y(t))$ la chaîne de Markov à temps continue et $E^{(\infty)}$ un espace d'états infini par

$$E^{(\infty)} = \{(i, j) : 0 \leq i \leq M_{RT}, j \geq 0\}, \quad (2.13)$$

où le symbole ∞ désigne un nombre infini d'appels BE. Notons également par $\nu(i)$ le temps de service d'appels BE et donc

$$\nu(i) = \mu_{BE} R_{BE}(i), \quad (2.14)$$

où $R_{BE}(i)$ est le taux de transmission total d'appels BE qui correspond à la capacité $C(i)$. Il donné par

$$R_{BE}(i) = \frac{C(i)}{1 - C(i)} \frac{W}{\Gamma E_{BE} / N_0}, \quad (2.15)$$

E_{BE} est l'énergie d'appels BE. La chaîne de Markov est ergodique si et seulement si le temps de service moyen des BE est supérieur à la charge d'appels BE, c'est à dire

$$\mu_{BE} E[R_{BE}(i)] > \lambda_{BE}, \quad (2.16)$$

$\mu_{BE} E[R_{BE}(i)] = \sum_{i=1}^{M_{RT}} \nu(i) q_i$. Le générateur infinitesimal de ce système est décrit par le

processus homogène de naissance et de mort indiqué par $\mathbf{Q}^{(\infty)}$ et donné par

$$\mathbf{Q}^{(\infty)} = \begin{pmatrix} \mathbf{B} & \mathbf{A}_0 & \mathbf{0} & \dots & \dots \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \ddots & \dots \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

où \mathbf{A}_0 , \mathbf{A}_1 , \mathbf{A}_2 et \mathbf{B} sont des matrices carrées d'ordres $M_{RT} + 1$ définies comme suit

$$\begin{aligned} \mathbf{A}_0 &= \text{diag}(\lambda_{BE}), \\ \mathbf{A}_2 &= \text{diag}(\nu(i); i = 0, 1, 2, \dots, M_{RT}), \\ \mathbf{A}_1[i, i+1] &= \lambda_{RT}; i = 0, 1, 2, \dots, M_{RT} - 1 \\ \mathbf{A}_1[i, i-1] &= i\mu_{RT}; i = 1, 2, \dots, M_{RT} \\ \mathbf{A}_1[i, i] &= -(\lambda_{BE} + \lambda_{RT} + i\mu_{RT} + \nu(i)), \\ &\quad i = 0, \dots, M_{RT} - 1, \\ \mathbf{A}_1[M_{RT}, M_{RT}] &= -(\lambda_{BE} + M_{RT}\mu_{RT} + \nu(M_{RT})), \\ \mathbf{B} &= \mathbf{A}_1 + \mathbf{A}_2. \end{aligned}$$

Nous supposons qu'en parallèle du cas infini, les appels BE demandent un délai moyen borné. Donc, le système traite un nombre fini d'appels BE pour assurer ce délai, noté M_{BE} . Pour cela, nous décrivons le système avec une chaîne de Markov homogène quasi-naissance et mort à temps continu notée $K'(t)$ où $K'(t) = (X(t), Y(t))$. Il en suit que l'espace d'états correspondant s'écrit

$$E^{(f)} = \{(i, j) : 0 \leq i \leq M_{RT}, 0 \leq j \leq M_{BE}\}, \quad (2.17)$$

le symbole f désigne le cas où le nombre d'appels BE est fini. Le générateur infinitésimal de ce système est

$$\mathbf{Q}^{(f)} = \begin{pmatrix} \mathbf{B} & \mathbf{A}_0 & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \dots & \ddots & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{A}_2 & \mathbf{F} \end{pmatrix},$$

où $\mathbf{F} = \mathbf{A}_1 + \mathbf{A}_0$, \mathbf{B} , \mathbf{A}_1 , \mathbf{A}_0 et \mathbf{A}_2 sont données dans le cas infini précédent.

2.2.3 Analyse du système

Dans cette section, nous analysons le système par l'approche de l'analyse spectrale avec le processus de naissance et de mort. Elle nous permettra de trouver la distribution stationnaire du système en fonction des valeurs et vecteurs propres de certaines

matrices finies de manière détaillée. Nous distinguons deux cas de figures : fini et infini.

Cas infini d'appels best effort : Soit $\pi^{(\infty)}$ le vecteur de probabilités d'états d'équilibres de la chaîne de Markov $(X(t), Y(t))$ et la probabilité d'avoir i appels de type RT et j appels de type BE est

$$\pi_{i,j}^{(\infty)} = \lim_{t \rightarrow \infty} Pr[X(t) = i, Y(t) = j]. \quad (2.18)$$

Sous la condition de stabilité du système, la matrice génératrice $Q^{(\infty)}$ est irréductible et apériodique. Par conséquent la distribution stationnaire de cette chaîne est la solution unique du système suivant

$$\pi^{(\infty)} \mathbf{Q}^{(\infty)} = \mathbf{0}, \quad (2.19)$$

$$\sum_{i=0}^{M_{RT}} \sum_{j=0}^{\infty} \pi_{i,j}^{(\infty)} = 1. \quad (2.20)$$

Notons que $\pi^{(\infty)}$ désigne un vecteur de blocs et il a la forme suivante

$$\pi^{(\infty)} = (\pi_0^{(\infty)}, \pi_1^{(\infty)}, \dots, \pi_j^{(\infty)}, \dots), \quad (2.21)$$

où $\pi_j^{(\infty)} = (\pi_{0,j}^{(\infty)}, \dots, \pi_{M_{RT},j}^{(\infty)})$. L'équation (2.19) peut se réécrire sous forme

$$\pi_0^{(\infty)} \mathbf{B} + \pi_1^{(\infty)} \mathbf{A}_2 = \mathbf{0}, \quad (2.22)$$

$$\pi_{n-1}^{(\infty)} \mathbf{A}_0 + \pi_n^{(\infty)} \mathbf{A}_1 + \pi_{n+1}^{(\infty)} \mathbf{A}_2 = \mathbf{0}, \quad n \geq 1. \quad (2.23)$$

Sous la condition de stabilité du système, la distribution stationnaire existe et est donnée par (Neuts, 1981) :

$$\pi_n^{(\infty)} = \pi_0^{(\infty)} \mathbf{R}^n, \quad n \geq 1, \quad (2.24)$$

où \mathbf{R} est une matrice carrée d'ordre $M_{RT} + 1$. Le rayon spectrale de la matrice \mathbf{R} est strictement inférieur à 1 et le vecteur $\pi_0^{(\infty)}$ est donné en combinant l'équation de frontière (2.24), la condition de normalisation (2.20) et les valeurs propres de la matrice \mathbf{R} . La matrice \mathbf{R} est la solution minimale non négative de l'équation suivante

$$\mathbf{A}_0 + \mathbf{R}\mathbf{A}_1 + \mathbf{R}^2\mathbf{A}_2 = \mathbf{0}. \quad (2.25)$$

Pour obtenir les valeurs propres de \mathbf{R} , nous définissons une matrice polynomial tridiagonale quadratique notée $\mathbf{T}(z)$ ($\forall z \in \mathbb{R}$)

$$\mathbf{T}(z) = \mathbf{A}_0 + z\mathbf{A}_1 + z^2\mathbf{A}_2. \quad (2.26)$$

Cas fini d'appels best effort : Puisque le générateur $\mathbf{Q}^{(f)}$ est apériodique et irréductible, alors il admet une distribution stationnaire de probabilités, unique, et solution de système des équations suivant

$$\pi^{(f)} \mathbf{Q}^{(f)} = \mathbf{0}, \quad (2.27)$$

$$\pi^{(f)} \mathbf{e}^{(f)} = \mathbf{1}, \quad (2.28)$$

où $e^{(f)}$ est le vecteur colonne des uns. Donc, nous obtenons le système suivant

$$\pi_0^{(f)} \mathbf{B} + \Pi_1^{(f)} \mathbf{A}_2 = 0, \quad (2.29)$$

$$\pi_{n-1}^{(f)} \mathbf{A}_0 + \pi_n^{(f)} \mathbf{A}_1 + \Pi_{n+1}^{(f)} \mathbf{A}_2 = 0, \quad (2.30)$$

$$\pi_{M_{BE}-1}^{(f)} \mathbf{A}_0 + \Pi_{M_{BE}}^{(f)} \mathbf{F} = 0, \quad (2.31)$$

pour $n = 1, \dots, M_{BE} - 1$.

Si $\lambda_{BE} \neq \sum_{i=0}^{M_{RT}} v(i)q_i$, alors le vecteur de probabilités $\pi^{(f)}$ d'appels BE s'écrit sous la forme

$$\pi_j^{(f)} = \bar{\mathbf{x}} \mathbf{R}^j + \bar{\mathbf{y}} \mathbf{S}^{M_{BE}-j}, \quad 0 \leq j \leq M_{BE}, \quad (2.32)$$

où \mathbf{R} et \mathbf{S} sont des matrices géométriques représentant les solutions minimales non négatives, respectivement, des équations suivantes

$$\mathbf{A}_0 + \mathbf{R} \mathbf{A}_1 + \mathbf{R}^2 \mathbf{A}_2 = \mathbf{0}, \quad (2.33)$$

$$\mathbf{S}^2 \mathbf{A}_0 + \mathbf{S} \mathbf{A}_1 + \mathbf{A}_2 = \mathbf{0}. \quad (2.34)$$

Pour obtenir les valeurs propres de \mathbf{R} et \mathbf{S} , nous définissons également deux matrices quadratiques polynomiales $\mathbf{T}(z)$ et $\mathbf{T}'(z)$ ($z \in \mathbb{R}$)

$$\mathbf{T}(z) = \mathbf{A}_0 + z \mathbf{A}_1 + z^2 \mathbf{A}_2, \quad (2.35)$$

$$\mathbf{T}'(z) = \mathbf{A}_2 + z \mathbf{A}_1 + z^2 \mathbf{A}_0. \quad (2.36)$$

2.2.4 Approche par l'analyse spectrale

Pour trouver les distributions stationnaires évoquées précédemment en fonction des valeurs et des vecteurs propres des matrices \mathbf{R} et \mathbf{S} , nous présentons quelques résultats préliminaires pour aboutir à ces distributions.

Lemme 1 1. Les racines non nulles du $\det[\mathbf{T}(z)]$ sont également les racines du $\det[(\mathbf{T}'(z))]$.

2. Les valeurs propres de la matrice \mathbf{R} (d'ordre m_i) sont les racines d'ordre supérieur ou égale à m_i du polynôme $\det[\mathbf{T}(z)]$.

3. Les vecteurs propres v_i de la matrice \mathbf{R} associés aux valeurs propres ϕ_i , sont les vecteurs qui annulent à gauche $\mathbf{T}(\phi_i)$.

4. Si $\phi_i \neq 0$ et $\frac{1}{\phi_i}$ une valeur propre de la matrice \mathbf{S} avec une multiplicité algébrique d'ordre m_i , alors ϕ_i est une racine du polynôme $\det[\mathbf{T}(z)]$ dont la multiplicité est d'ordre au moins égale à m_i . Si v_i est un vecteur propre de la matrice \mathbf{S} correspondant à la valeur propre $\frac{1}{\phi_i}$, alors v_i est un vecteur qui annule à gauche $\mathbf{T}(\phi_i)$.

Démonstration du Lemme 1 1. Nous pouvons écrire l'équation $\mathbf{T}(z)$ pour $z \neq 0$ comme $\mathbf{T}(z) = z^2 \mathbf{T}'(1/z)$. Nous remarquons que les racines du polynôme $\det[\mathbf{T}'(z)]$ sont exactement les racines du polynôme $\det[\mathbf{T}(z)]$.

2. A partir de l'équation (2.25) nous avons $\mathbf{A}_0 = -\mathbf{R}(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2)$. En remplaçant celle-ci dans (2.26) nous obtenons

$$\mathbf{T}(z) = -\mathbf{R}(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2) + z(\mathbf{A}_1 + z\mathbf{A}_2) + z\mathbf{R}\mathbf{A}_2 - z\mathbf{R}\mathbf{A}_2. \quad (2.37)$$

Donc, nous avons

$$\mathbf{T}(z) = (z\mathbf{I} - \mathbf{R})(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2 + z\mathbf{A}_2). \quad (2.38)$$

Le déterminant de la matrice $\mathbf{T}(z)$ est décomposé sous la forme

$$\det[\mathbf{T}(z)] = \det(z\mathbf{I} - \mathbf{R})\det(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2 + z\mathbf{A}_2). \quad (2.39)$$

De plus, $P_{\mathbf{R}}(z) = \det(z\mathbf{I} - \mathbf{R})$ est un polynôme caractéristique de la matrice \mathbf{R} et ϕ_i sa valeur propre d'ordre m_i . Nous obtenons la forme suivante pour $z = \phi_i$

$$\det[\mathbf{T}(\phi_i)] = P_{\mathbf{R}}(\phi_i)\det(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2 + \phi_i\mathbf{A}_2). \quad (2.40)$$

Puisque la valeur propre ϕ_i est une racine de la matrice $\det(\mathbf{A}_1 + \mathbf{R}\mathbf{A}_2 + z\mathbf{A}_2)$ d'ordre n_i , alors

$$\det[\mathbf{T}(z)] = (z - \phi_i)^{m_i+n_i}g(z), \quad (2.41)$$

où $g(z)$ est un polynôme de degré $M_{\mathbf{RT}} - m_i - n_i + 1$. Par conséquent $\det[\mathbf{T}(\phi_i)] = 0$ où ϕ_i est une racine d'ordre $m_i + n_i$.

3. Soit v_i le vecteur propre associé à la valeur propre ϕ_i . Nous multiplions l'équation (2.26) par le vecteur propre v_i et utilisant l'équation (2.33), nous trouvons

$$\begin{aligned} v_i\mathbf{T}(\phi_i) &= v_i\mathbf{A}_0 + \phi_i v_i\mathbf{A}_1 + v_i\phi_i^2\mathbf{A}_2 \\ &= v_i\mathbf{A}_0 + v_i\mathbf{R}\mathbf{A}_1 + v_i\mathbf{R}^2\mathbf{A}_2 \\ &= v_i(\mathbf{A}_0 + \mathbf{R}\mathbf{A}_1 + \mathbf{R}^2\mathbf{A}_2) \\ &= \mathbf{0}. \end{aligned}$$

Ce qui complète la démonstration.

4. Nous remarquons que $\mathbf{T}'(z) = (z\mathbf{I} - \mathbf{S})((z\mathbf{I} + \mathbf{S})\mathbf{A}_0 + \mathbf{A}_1)$. Soit $\frac{1}{\phi_i}$ une valeur propre de \mathbf{S} , alors $\frac{1}{\phi_i}$ est une racine de $\det[\mathbf{T}'(z)]$ de multiplicité au moins égale m_i . Nous avons également

$$v_i\mathbf{T}(\phi_i) = v_i\phi_i^2\mathbf{T}'(1/\phi_i) = \phi_i^2\left(\frac{1}{\phi_i}v_i - v_i\mathbf{S}\right)\left(\left(\frac{1}{\phi_i}\mathbf{I} + \mathbf{S}\right)\mathbf{A}_0 + \mathbf{A}_1\right) = \mathbf{0}.$$

La dernière equation est due à la relation $v_i\mathbf{S} = \frac{1}{\phi_i}v_i$.

- Lemme 2**
1. La matrice $\mathbf{T}(z)$ a une valeur propre $\lambda_{\mathbf{BE}}$ de multiplicité $M_{\mathbf{RT}} + 1$ pour $z = 0$.
 2. Si $z \neq 0$, la matrice $\mathbf{T}(z)$ a $M_{\mathbf{RT}} + 1$ valeurs propres différentes.
 3. Le déterminant de la matrice $\mathbf{T}(z)$ peut s'écrire sous la forme

$$\det[\mathbf{T}(z)] = (1 - z)d(z), \quad (2.42)$$

où $d(z)$ est un polynôme de degré $2M_{\mathbf{RT}} + 1$.

Démonstration du Lemme 2 1. Nous posons $\mathbf{T}(0) = \mathbf{A}_0$ et pour un réel λ , nous écrivons le polynôme caractéristique de $\mathbf{T}(0)$ sous la forme

$$P_{\mathbf{T}(0)}(\lambda) = P_{\mathbf{A}_0}(\lambda) = (\lambda - \lambda_{BE})^{M_{RT}+1}. \quad (2.43)$$

Par conséquent $\lambda = \lambda_{BE}$ est une valeur propre d'ordre $M_{RT} + 1$ de la matrice $\mathbf{T}(z)$ à $z = 0$.

2. Pour $z \in \mathbb{R}^*$. Soit $\mathbf{D} = \text{diag}(d_i)$ une matrice diagonale telle que la matrice $\mathbf{DT}(z)\mathbf{D}^{-1}$ est symétrique. Les éléments de la matrice \mathbf{D} sont obtenus par la relation $\frac{d_i}{d_0} = \sqrt{\left(\frac{\lambda_{RT}}{\mu_{RT}}\right)^i \frac{1}{i!}}$ pour $i = 0, \dots, M_{RT}$. On peut vérifier facilement que les éléments de la matrice $\mathbf{E}(z) = \mathbf{DT}(z)\mathbf{D}^{-1}$ peuvent s'écrire comme suit

$$\begin{aligned} \mathbf{E}(z)[i, i] &= k(i, z), \quad \text{for } i = 0, \dots, M_{RT}, \\ \mathbf{E}(z)[i, i+1] &= z\sqrt{i\lambda_{RT}\mu_{RT}}, \\ &\quad i = 0, \dots, M_{RT} - 1, \\ \mathbf{E}(z)[i, i-1] &= z\sqrt{(i-1)\lambda_{RT}\mu_{RT}}, \\ &\quad i = 1, \dots, M_{RT}, \end{aligned}$$

où

$$\begin{aligned} k(i, z) &= \lambda_{BE} - z(\lambda_{RT} + \lambda_{BE} + i\mu_{RT} + \nu(i)) + z^2\nu(i), \\ k(M_{RT}, z) &= \lambda_{BE} - z(\lambda_{BE} + M_{RT}\mu_{RT} + \nu(M_{RT})) + z^2\nu(M_{RT}), \end{aligned}$$

pour $i = 0, 1, \dots, M_{RT} - 1$. La matrice $\mathbf{E}(z)$ est réelle et symétrique d'ordre $M_{RT} + 1$, alors $\mathbf{E}(z)$ a $M_{RT} + 1$ vecteurs propres associés aux $M_{RT} + 1$ valeurs propres réelles (voir (Burden et Faires, 1997)). Ces vecteurs propres forment une base orthonormale. En plus elles constituent une famille libre ce qui montre que la matrice $\mathbf{E}(z)$ a $M_{RT} + 1$ valeurs propres différentes.

3. Le déterminant de la matrice $\mathbf{T}(z)$ peut se réécrire sous la forme (2.44).

$$\det[\mathbf{T}(z)] = \begin{vmatrix} k(0, z) & z\lambda_{RT} & 0 & \dots & 0 \\ z\mu_{RT} & k(1, z) & z\lambda_{RT} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & z\lambda_{RT} \\ 0 & \dots & 0 & zM_{RT}\mu_{RT} & k(M_{RT}, z) \end{vmatrix}. \quad (2.44)$$

En remplaçant la dernière colonne de la matrice $\det[\mathbf{T}(z)]$ par la somme de toutes les colonnes, nous trouvons

$$\det[\mathbf{T}(z)] = (1 - z)d(z), \quad (2.45)$$

où $d(z)$ est de la forme

$$d(z) = \begin{vmatrix} k(0, z) & z\lambda_{RT} & 0 & \dots & 0 & (\lambda_{BE} - z\nu(0)) \\ z\mu_{RT} & k(1, z) & z\lambda_{RT} & \ddots & \vdots & (\lambda_{BE} - z\nu(1)) \\ 0 & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & & z\lambda_{RT} & \vdots \\ \vdots & \vdots & 0 & z(M_{RT} - 1)\mu_{RT} & k(M_{RT} - 1, z) & (\lambda_{BE} - z\nu(M_{RT} - 1)) \\ 0 & 0 & \dots & 0 & zM_{RT}\mu_{RT} & (\lambda_{BE} - z\nu(M_{RT})). \end{vmatrix}.$$

Puisque les matrices $\mathbf{E}(z)$ et $\mathbf{T}(z)$ sont similaires, alors elles ont les mêmes valeurs propres. Ces valeurs propres sont $\theta_0(z), \dots, \theta_{M_{RT}}(z)$. Soit $P_{\mathbf{T}(z)}(\lambda) = \det(\lambda \mathbf{I} - \mathbf{T}(z))$ le polynôme caractéristique de la matrice polynômiale $\mathbf{T}(z)$ et un réel λ . Nous appliquons le lemme 2 et nous avons

$$P_{\mathbf{T}(z)}(\lambda) = (\lambda - \theta_0(z))(\lambda - \theta_1(z)) \dots (\lambda - \theta_{M_{RT}}(z)). \quad (2.46)$$

Et la trace de $\mathbf{T}(z)$ est donnée par

$$Tr[\mathbf{T}(z)] = (M_{RT} + 1)\lambda_{BE} - M_{RT}\lambda_{RT} \sum_{i=0}^{M_{RT}} (\lambda_{BE} + i\mu_{RT} + \nu(i))z + \sum_{i=0}^{M_{RT}} \nu(i)z^2.$$

Alors

$$P_{\mathbf{T}(z)}(\lambda) = \alpha_0(z) + \alpha_1(z)\lambda + \dots + \alpha_{M_{RT}-1}(z)\lambda^{M_{RT}-1} + \lambda^{M_{RT}},$$

où $\alpha_i(z)$ est une fonction polynômiale de degré 2 et la trace de la matrice $Tr[\mathbf{T}(z)]$ est (Burden et Faires, 1997)

$$Tr[\mathbf{T}(z)] = \alpha_{M_{RT}-1}(z)$$

et la fonction $\alpha_{M_{RT}-1}(z)$ est aussi continue. Donc $\alpha_{M_{RT}-1}(z) = \sum_{i=0}^{M_{RT}} \theta_i(z)$ et nous déduisons la continuité des valeurs propres $\theta_i(z) (\forall z \in \mathbb{R})$. Nous avons $P_{\mathbf{T}(z)}(0) = \det[\mathbf{T}(z)] = \alpha_0(z)$ et

$$\det[\mathbf{T}(z)] = \theta_0(z)\theta_1(z) \dots \theta_{M_{RT}}(z). \quad (2.47)$$

Nous avons vu que les fonctions $\theta_k(z)$ sont continues pour tout z . En plus, nous pouvons écrire les valeurs propres selon le lemme 2 sous la forme

1. Si $z \neq 0$ alors nous avons

$$\theta_0(z) < \theta_1(z) < \dots < \theta_{M_{RT}}(z). \quad (2.48)$$

2. Si $z = 0$ alors nous avons

$$\theta_0(0) = \theta_1(0) = \dots = \theta_{M_{RT}}(0) = \{\lambda_{BE}\}. \quad (2.49)$$

Lemme 3 1. Toutes les valeurs propres $\theta_i(z)$ sont négatifs à $z = 1$, cela signifie que $\theta_i(1) \leq 0$ pour $i = 0, 1, \dots, M_{RT}$.

2. Pour $z \in]0, 1[$, les équations des valeurs propres en fonction de z , $\theta_i(z) = 0$, ont au moins une solution, ce qui signifie que $\exists \psi_i \in]0, 1[$ tel que $\theta_i(\psi_i) = 0$ pour $i = 0, 1, \dots, M_{RT}$.

Démonstration du Lemme 3 1. Pour faciliter la notation, nous posons $\mathbf{Q}_{RT} = \mathbf{T}(1)$ où \mathbf{Q}_{RT} est une matrice carrée d'ordre $M_{RT} + 1$ définie par $\mathbf{Q}_{RT} = (a_{i,j})_{0 \leq i,j \leq M_{RT}}$. Nous constatons que les éléments diagonaux de cette matrice sont tous négatives. En appliquant le théorème de Gersgorin détaillé dans (Burden et Faires, 1997), nous disons que chaque valeur propre $\theta_i(1)$ de la matrice $\mathbf{T}(1)$ appartient au moins à un disque de type $\mathbf{D}(a_{i,i}, r_i)$ dont le centre est

$$a_{i,i} = -(\lambda_{RT} + i\mu_{RT}) \quad (2.50)$$

et le rayon est

$$r_i = \lambda_{RT} + i\mu_{RT}, \quad i = 0, \dots, M_{RT}. \quad (2.51)$$

D'où la démonstration est terminée par $|\theta_i(1) - a_{i,i}| \leq r_i$, ce qui signifie que nous aurons $\theta_i(1) \leq r_i + a_{i,i} = 0$.

2. Nous avons $\theta_i(0) > 0$, $\theta_i(1) < 0$ et la continuité de ces fonctions, ce qui montre qu'il existe au moins une solution ψ_i dans l'intervalle $]0, 1[$ tel que $\theta_i(\psi_i) = 0$.

Théorème 1 Le polynôme $\det[\mathbf{T}(z)]$ possède au maximum $2M_{RT} + 2$ racines non nulles localisées selon les trois cas :

1. si $\lambda_{BE} < \sum_{i=0}^{M_{RT}} \nu(i)q_i$, alors nous avons une racine à $z = 1$, $M_{RT} + 1$ racines dans le disque unité M_{RT} à l'extérieur du disque unité ;
2. si $\lambda_{BE} = \sum_{i=0}^{M_{RT}} \nu(i)q_i$, nous avons une racine d'ordre 2 à $z = 1$, $M_{RT} + 1$ dans le disque unité et $M_{RT} - 1$ à l'extérieur du disque unité ;
3. si $\lambda_{BE} > \sum_{i=0}^{M_{RT}} \nu(i)q_i$, nous avons M_{RT} racines dans le disque unité, une à $z = 1$ et $M_{RT} + 1$ à l'extérieur du disque unité.

Démonstration du Théorème 1 1. Nous décomposons le polynôme $\det[\mathbf{T}(z)]$ en utilisant le lemme 2 sous la forme

$$\det[\mathbf{T}(z)] = (1 - z)d(z), \quad (2.52)$$

où $d(z)$ est un polynôme de degré $2M_{RT} + 1$. Nous remarquons que le polynôme $\det[\mathbf{T}(z)]$ possède au maximum $2M_{RT} + 2$ racines.

2. Si $\lambda_{BE} < \sum_{i=0}^{M_{RT}} \nu(i)q_i$, Alors le rayon spectral de la matrice R est strictement inférieur à 1, c'est à dire $sp(R) < 1$, voir Théorème 3.1.1 (Neuts, 1981). A partir du lemme 6 nous montrons que toutes les valeurs propres de la matrice R sont aussi des racines du polynôme $\det[\mathbf{T}(z)]$. Par conséquent, il y a $M_{RT} + 1$ racines du polynôme $\det[\mathbf{T}(z)]$ dans le disque unité pour $|z| < 1$. Ces valeurs propres sont réelles et strictement positives. Quand $z = 1$, il y a une racine d'ordre 1 du $\det[\mathbf{T}(z)]$ selon les lemmes 2 et 3. L'inverse de ces racines est valeur propre de la matrice S et ces valeurs sont toutes des racines du $\det[\mathbf{T}(z)]$ pour $|z| > 1$.

3. Nous calculons d'abord $d(1)$ par l'expression

$$d(1) = (-1)^{M_{RT}+1} (\mu_{RT})^{M_{RT}} \sum_{i=0}^{M_{RT}} (\lambda_{BE} - \nu(i)) \left(\frac{\lambda_{RT}}{\mu_{RT}}\right)^i \frac{M_{RT}!}{i!}.$$

Nous remplaçons dans l'expression ci-dessus $\left(\frac{\lambda_{RT}}{\mu_{RT}}\right)^i$ par $\frac{q_i}{q_0} i!$ et retrouvons

$$\begin{aligned} d(1) &= (-1)^{M_{RT}+1} (\mu_{RT})^{M_{RT}} \frac{M_{RT}!}{q_0} \sum_{i=0}^{M_{RT}} (\lambda_{BE} q_i - \nu(i) q_i), \\ d(1) &= (-1)^{M_{RT}+1} (\lambda_{BE} - \omega) \left(\frac{M_{RT}!}{q_0}\right) (\mu_{RT})^{M_{RT}}, \end{aligned} \tag{2.53}$$

où $\omega = \sum_{i=0}^{M_{RT}} \nu(i) q_i$. Si $\lambda_{BE} = \omega$ alors dans l'expression (2.53) nous trouvons $d(1) = 0$.

Ce qui montre qu'il y a une autre racine quand $z = 1$, d'où il y a une racine d'ordre 2 à $z = 1$ et $M_{RT} + 1$ racines dans le disque unité (voir les lemmes 2 et 3).

4. La preuve est similaire au troisième cas.

Jusqu'à maintenant, nous avons trouvé les valeurs et les vecteurs propres des matrices \mathbf{R} et \mathbf{S} nécessaires pour calculer la distribution stationnaire du nombre d'appels RT et BE. Mais d'autres coefficients, restent inconnus pour exprimer correctement cette distribution en fonction de ces valeurs et vecteurs propres. Ces coefficients font l'objet de ce qui suit.

2.2.5 Distribution stationnaire

Cas infini d'appels best effort : Nous écrivons la probabilité d'état d'équilibre sous la forme spectrale

$$\pi_j^{(\infty)} = \sum_{i=0}^{M_{RT}} \beta_i \phi_i^j v_i, \quad j = 0, 1, 2, \dots \tag{2.54}$$

D'après le résultat de l'analyse précédente, la matrice \mathbf{R} a $M_{RT} + 1$ valeurs propres simples qui sont $\phi_0, \dots, \phi_{M_{RT}}$. Il existe une matrice \mathbf{V} inversible composée par les vecteurs propres de la matrice \mathbf{R} tel que

$$\mathbf{R} = \mathbf{V}^{-1} \Phi \mathbf{V},$$

où

$$\begin{aligned} \Phi &= \text{diag}(\phi_0, \dots, \phi_{M_{RT}}), \\ \mathbf{V} &= \begin{pmatrix} v_0 \\ \vdots \\ v_{M_{RT}} \end{pmatrix}. \end{aligned}$$

L'équation (2.54) peut s'écrire sous la forme matricielle

$$\pi_j^{(\infty)} = \pi_0^{(\infty)} \mathbf{V}^{-1} \Phi^j \mathbf{V}. \quad (2.55)$$

Alors, le résultat du produit $\pi_0^{(\infty)} \mathbf{V}^{-1}$ est noté $\bar{\beta}$. Ce qui nous a permis de réécrire l'équation (2.54) comme suit

$$\pi_j^{(\infty)} = \bar{\beta} \Phi^j \mathbf{V}, \quad j = 0, 1, \dots, \quad (2.56)$$

où

$$\bar{\beta} = (\beta_0, \beta_1, \dots, \beta_{M_{RT}}).$$

Cependant, la condition de normalisation devient sous la forme suivante

$$\sum_{j=0}^{\infty} \pi_j^{(\infty)} \bar{\mathbf{1}} = \sum_{i=0}^{M_{RT}} \frac{\beta_i}{1 - \phi_i} v_i \bar{\mathbf{1}} = 1, \quad (2.57)$$

où $\bar{\mathbf{1}}$ est un vecteur colonne d'ordre $M_{RT} + 1$. Nous combinons l'expression (2.56), les équations de frontière (2.22) et la condition de la normalisation (2.57), pour trouver les seules inconnues β_k comme des solutions du système linéaire suivant :

$$\bar{\beta} [\mathbf{V}\mathbf{B} + \Phi \mathbf{V}\mathbf{A}_2, \Psi \mathbf{V}\bar{\mathbf{1}}] = [\mathbf{0}, 1],$$

où

$$\Psi = \text{diag}\left(\frac{1}{1 - \phi_i}; i = 0, 1, \dots, M_{RT}\right).$$

Cas fini d'appels best effort : De la même façon que dans le cas précédent, en utilisant la décomposition de Jordan, nous écrivons les deux matrices \mathbf{R} et \mathbf{S} comme suit

$$\mathbf{S} = \mathbf{W}^{-1} \Omega \mathbf{W},$$

où

$$\Omega = \text{diag}\left(\frac{1}{\phi_i}; i = M_{RT} + 1, \dots, 2M_{RT} + 2\right),$$

$$\mathbf{W} = \begin{pmatrix} v_{M_{RT}+1} \\ \vdots \\ v_{2M_{RT}+2} \end{pmatrix}.$$

Si $\lambda_{BE} \neq \sum_{i=0}^{M_{RT}} \nu(i) q_i$ et les valeurs propres de la matrice \mathbf{R} sont non nulles et simples, alors nous pouvons réécrire l'équation (2.32) pour $j = 0, 1, \dots, M_{BE}$ comme suit

$$\pi_j^{(f)} = \bar{\mathbf{x}} \mathbf{V}^{-1} \Psi^j \mathbf{V} + \bar{\mathbf{y}} \mathbf{W}^{-1} \Omega^{M_{BE}-j} \mathbf{W}. \quad (2.58)$$

La forme finale de la distribution stationnaire dans ce cas est

$$\pi_j^{(f)} = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^j v_i + \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i \left(\frac{1}{\phi_i}\right)^{M_{BE}-j} v_i. \quad (2.59)$$

Les seules inconnues maintenant sont les coefficients β_i . Ces coefficients seront obtenus comme solution du système des équations formées en combinant l'expression (2.59) avec les équations de frontières (2.29), (2.31) et l'équation de normalisation (2.28), comme suit

$$(\beta_0, \dots, \beta_{2M_{RT}+2}) \times \begin{bmatrix} \mathbf{VB} + \Phi \mathbf{VA}_2 & \Phi^{(M_{BE}-1)} \mathbf{VA}_0 + \Phi^{M_{BE}} \mathbf{VF} \\ \Omega^{M_{BE}} \mathbf{WB} + \Omega^{(M_{BE}-1)} \mathbf{WA}_2 & \Omega \mathbf{WA}_0 + \mathbf{WF} \end{bmatrix} = (\mathbf{0}, \mathbf{0}).$$

Résumé des résultats : Maintenant, nous résumons les résultats trouvés par notre analyse. Les distributions de probabilités pour le cas fini donnée dans (2.24) et le cas infini dans (2.32) sont réécrites sous les formes d'analyse spectrale comme suit :

1. Cas infini d'appels BE

$$\pi_j^{(\infty)} = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^j v_i, \quad j = 0, 1, 2, \dots, \quad (2.60)$$

où $\phi_0, \dots, \phi_{M_{RT}}$ sont les valeurs propres associées aux vecteurs propres $v_0, \dots, v_{M_{RT}}$ de la matrice \mathbf{R} , c'est à dire $v_i \mathbf{R} = \phi_i v_i$ et $\beta_0, \dots, \beta_{M_{RT}}$ sont des coefficients réels.

2. Cas fini d'appels BE

$$\pi_j^{(f)} = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^j v_i + \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i \left(\frac{1}{\phi_i}\right)^{M_{BE}-j} v_i, \quad (2.61)$$

où $j = 0, 1, \dots, M_{BE}$; les valeurs $\phi_0, \dots, \phi_{M_{RT}}$ sont les valeurs propres associées aux vecteurs propres $v_0, \dots, v_{M_{RT}}$ de la matrice \mathbf{R} ; les valeurs $1/\phi_{M_{RT}+1}, \dots, 1/\phi_{2M_{RT}+2}$ sont des valeurs propres associées aux vecteurs propres gauches $v_{M_{RT}+1}, \dots, v_{2M_{RT}+2}$ de la matrice \mathbf{S} et $\beta_0, \dots, \beta_{2M_{RT}+2}$ sont des coefficients réels pour les deux vecteurs : \bar{x} et \bar{y} .

2.2.6 Métriques de performances

Après avoir trouvé les formes explicites de la distribution stationnaire dans les deux cas étudiés précédemment, nous fournissons quelques métriques de performances du système, comme la probabilité marginale, le temps de séjour et le nombre moyen marginale d'appels BE.

Cas infini d'appels best effort : Puisque les appels BE dépendent de ceux en RT dans le système, la probabilité marginale des BE est une métrique de mesure importante qui nous permet d'évaluer les performances du système en terme d'appels BE par rapport

aux différents configurations possibles de ceux en RT. Alors la probabilité marginale $P_j^{(\infty)}$ est donnée par

$$P_j^{(\infty)} = \sum_{i=0}^{M_{RT}} \beta_i \phi_i^j v_i \bar{1}. \quad (2.62)$$

Ansi, le nombre marginal moyen d'appels BE et le temps de séjour de ces appels sont illustrés respectivement par :

$$E_{BE}^{(\infty)} = \sum_{i=0}^{M_{RT}} \beta_i \frac{\phi_i}{(1 - \phi_i)^2} v_i \bar{1}, \quad (2.63)$$

$$T_{BE}^{(\infty)} = \frac{E_{BE}^{(\infty)}}{\lambda_{BE}}. \quad (2.64)$$

Nous constatons que ces métriques citées ci-dessus sont indépendantes de nombre d'appels BE dans le système. Ce qui permet de simplifier la complexité de calcul des différentes métriques de performances du système.

Cas fini d'appels best effort : La probabilité marginale dans ce cas est définie comme suit

$$P_j^{(f)} = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^j v_i \bar{1} + \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i \left(\frac{1}{\phi_i}\right)^{M_{BE}-j} v_i \bar{1}. \quad (2.65)$$

Nous déduisons la probabilité de blocage d'un nouvel appel de type BE qui est calculée dans la formule ci-dessus quand le nombre de BE est exactement égale M_{BE} . D'où

$$P_{M_{BE}}^{(f)} = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^{M_{BE}} v_i \bar{1} + \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i v_i \bar{1}. \quad (2.66)$$

Le nombre marginal moyen et le temps de séjour en utilisant la formule de Little pour les appels BE sont respectivement :

$$E_{BE}^{(f)} = \sum_{j=0}^{M_{BE}} j P_j^{(f)}, \quad (2.67)$$

$$T_{BE}^{(f)} = \frac{E_{BE}^{(f)}}{\lambda_{BE}(1 - P_{M_{BE}}^{(f)})}. \quad (2.68)$$

2.2.7 Cas multi cellules

Dans cette section nous étendons notre analyse au cas multi cellules en introduisant une approximation aux interférences inter cellules (I_{other}). Nous supposons que toutes les cellules sont statistiquement identiques et homogènes. Ainsi, le nombre de classes de service dans la cellule est un nombre aléatoire. Il dépend du taux d'arrivées d'appels et de leurs durées de service. Ce n'est pas raisonnable de considérer que la relation (2.3) soit valide tout le temps. Cependant, nous supposons qu'elle est vraie par estimation

et particulièrement dans un réseau de densité forte, nous avons $E[I_{other}] = gE[I_{own}]$. En plus, la puissance minimale (P_i) donnée par (2.4) devient

$$P_i = \frac{N\Delta_i}{1 - X_{RT}\Delta_{RT} - X_{BE}\Delta_{BE} - gE[I_{own}]} \quad (2.69)$$

où

$$\begin{aligned} E[I_{own}] &= E[X_{RT}\Delta_{RT}] + E[X_{BE}\Delta_{BE}] \\ &= \sum_{i=0}^{M_{RT}} iq_i\Delta_{RT} + (1 - \mathbf{\Pi}_0^{BE}\mathbf{e}) \times \\ &\quad \sum_{i=0}^{M_{RT}} q_iC(i) \end{aligned} \quad (2.70)$$

où q_i est donné par (2.11), $\mathbf{\Pi}_0^{BE}$ est obtenu par la relation (2.56) pour le cas infini et à partir de la relation (2.59) pour le cas fini, et \mathbf{e} est un vecteur colonne de même dimension. Le second terme de l'équation (2.70) est donné par le fait que les appels BE utilisent tous la capacité restante quand il y a au moins un appel d'eux. La nouvelle capacité devient inconnue et dépend de la distribution stationnaire du système, elle s'écrit sous la forme $\Theta = 1 - gE[I_{own}]$. Dans un but d'obtenir la capacité totale du système nous utilisons l'algorithme suivant

Algorithm 1 : Convergence de la capacité totale

- 1: Initialiser la capacité de la cellule par $\Theta_{old} = 1 - gI_0$, où I_0 est la valeur initiale des interférences $E[I_{own}]$.
 - 2: Calculer la distribution de probabilité $\mathbf{\Pi}_0^{BE}$ par (2.56) pour le cas infini et par(2.59) dans le cas fini.
 - 3: Calculer l'interférence total estimée ($E[I_{own}]$) due aux appels RT et BE par l'équation (2.70).
 - 4: Dédire la capacité totale notée par $\Theta_{new} = 1 - gE[I_{own}]$.
 - 5: Vérifier la convergence de la capacité entre l'ancienne et la nouvelle valeur, c'est à dire si $|\Theta_{new} - \Theta_{old}| < \epsilon_0$, où ϵ_0 est un nombre très petit et strictement positive, alors la nouvelle capacité est utilisée pour l'analyse du système. Sinon, passer à l'étape 2 avec la nouvelle capacité. Les itérations sont poursuivies jusqu'à atteindre la convergence de la capacité.
-

2.2.8 Résultats numériques

Nous utilisons le codec vocal AMR pour les appels RT, qui propose huit taux de transmissions allant de 4,75 à 12,2 Kbps. Ces taux peuvent changer dynamiquement toutes les 20 ms. Le taux total de chargement normalisé et demandé par les appels RT est le suivant $R_T = |\Theta_\epsilon / \Delta_{RT}^m| R_{RT}^m$. La valeur du paramètre ϵ est égale à 10^{-5} . Ainsi, les taux de charges normalisés pour RT et BE sont respectivement : $\bar{\rho}_{BE} = \frac{\lambda_{BE}}{\mu_{BE}} \frac{1}{R_T}$ et

$\bar{\rho}_{RT} = \rho_{RT} \frac{R_{RT}^m}{R_T}$. Nos applications numériques sont basées sur les valeurs des paramètres standards du système WCDMA et sont résumées dans le tableau 2.1.

Energie du signal par bit divisée par la densité du bruit pour le trafic RT, E_{RT}/N_0	4.1 dB
Energie du signal par bit divisée par la densité du bruit pour le trafic NRT, E_{BE}/N_0	3.1 dB
Débit minimum d'appels RT, R_{RT}^m	4.27 Kbps
Débit maximum d'appels RT, R_{RT}^r	12.2 Kbps
Charge total d'appels RT, R_T	38 Kbps
Débit chip WCDMA, W	3.84 Mcps
Taux d'arrivée d'appels BE, λ_{BE}	10 appels/s
Durée moyenne d'une session BE, $1/\mu_{BE}$	1/5 Mbits
Taux de charge normalisé d'appels BE, $\bar{\rho}_{BE}$	0.5
Taux d'arrivée d'appels RT, λ_{RT}	0.3 appels/s
Durée moyenne de service RT, $1/\mu_{RT}$	10 s
Taux de charge normalisé d'appels RT, $\bar{\rho}_{RT}$	0.3

TABLE 2.1 – Valeurs numériques pour WCDMA (Holma et Toskala, 2001; Hegde et Altman, 2003)

La figure 2.1 expose la probabilité de blocage d'appels RT en fonction de la proportion de la capacité réservée aux appels BE L_{BE} . Les résultats de cette figure sont obtenus pour différentes charges du réseau par les appels RT. Selon cette figure nous remarquons bien que lorsque L_{BE} augmente la probabilité augmente, ceci est justifié par le fait qu'il reste moins de ressources pour les appels RT. En plus, une remarque importante s'ajoute c'est que le blocage croît rapidement pour les valeurs de L_{BE} supérieures à 0.55, car il y a moins de ressources restantes pour les appels RT et surtout dans la forte densité de charge du réseau $\bar{\rho}_{RT} = 0.8$.

Nous gardons les taux de charges constants et varions les temps de services d'appels RT, $1/\mu_{RT}$, dans la plage de valeurs $\{100, 10, 1, 0.01\}$. Nous évaluons les performances du système de manière numérique en fonction de la capacité réservée aux appels BE. En effet, la figure 2.2 présente le temps du séjour des BE pour différentes valeurs du temps de service d'appels RT. Nous constatons que le temps de séjour dans ces différentes configurations décroît rapidement quand les valeurs du seuil L_{BE} sont faibles et devient constant juste après la valeur 0.005 ms. Par conséquent, un résultat très important de cette figure est que pour une capacité importante réservée aux appels BE (plus de 50%), leur temps de séjour devient insensible à la durée de service d'appels RT. Donc, l'impact d'appels RT sur ceux en BE devient négligeable.

Nous essayons dans ce paragraphe de définir un contrôle d'admission d'appels BE. Supposons que l'opérateur ou le fournisseur de service (Internet Service Provider) veut imposer une qualité de service en terme du temps de séjour pour les appels BE. Ce temps de séjour ou un seuil de séjour est déterminé via la figure 2.3. Cette figure représente le temps de séjour moyen d'appels BE en fonction de la proportion de la capacité

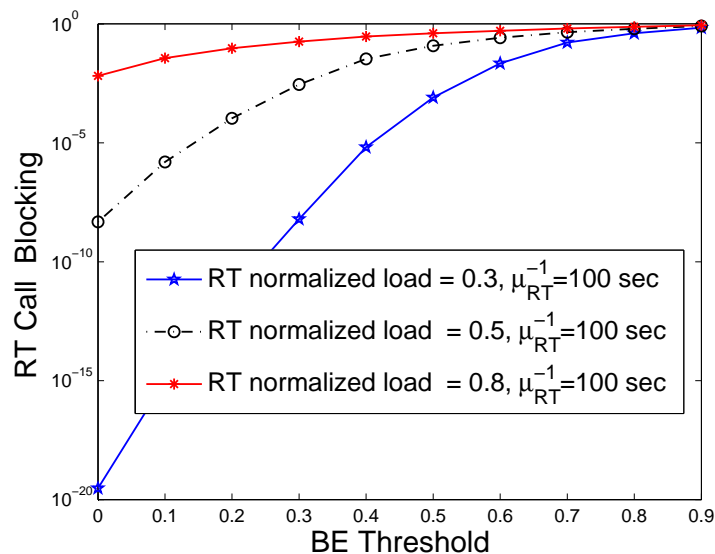


FIGURE 2.1 – Probabilité de blocage d'appels RT en fonction du seuil de la capacité réservée aux appels BE, pour $\bar{\rho}_{RT} = 0.3$ et $\mu_{RT} = 0.1$.

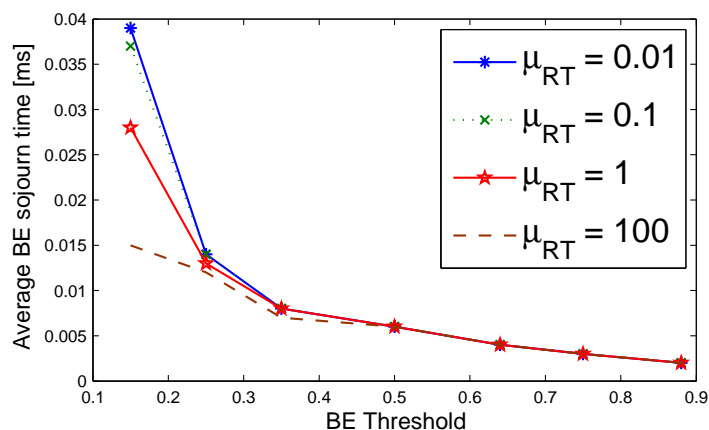


FIGURE 2.2 – Temps de séjour d'appels BE en fonction du seuil de la capacité réservée aux appels BE pour différents taux de service d'appels RT, $\mu_{BE} = 5$, $\lambda_{BE} = 10$ et $\bar{\rho}_{RT} = 0.5$.

réservée aux appels BE (L_{BE}). Remarquons ici que le temps de séjour prévu diminue remarquablement avec l'augmentation du seuil L_{BE} . Alors une petite réservation de ressources pour les appels BE implique une longue durée de séjour de ces appels. Notons que ce résultat confirme celui trouvé dans (Altman, 2002). Maintenant, nous supposons que l'opérateur via les statistiques historiques d'utilisation du canal par les appels RT et BE, connaît comment il peut partager les ressources entre les deux classes d'appels. La qualité de service visée par l'opérateur est que le temps de séjour soit borné. Par exemple, s'il réserve aux appels BE la proportion $L_{BE} = 0.3$, la qualité de service est que le temps de séjour d'appels BE soit inférieur à 0.015 ms.

La figure 2.4 illustre le taux maximum d'arrivée (λ_{BE}^*) d'appels BE en fonction du seuil du temps de séjour de ces appels. Nous constatons que le taux maximum d'appels BE varie conjointement avec le temps de séjour, quand ce temps est faible (moins de 0.09 ms), le taux correspondant est grand. Aussi quand l'opérateur propose un seuil de temps de séjour supérieur à 0.1 ms , le taux maximum d'arrivée devient légèrement constant. Ce qui nous permet de fixer une qualité de service pour les appels BE comme le temps de leurs séjour ne doit pas dépasser le seuil 0.15 ms .

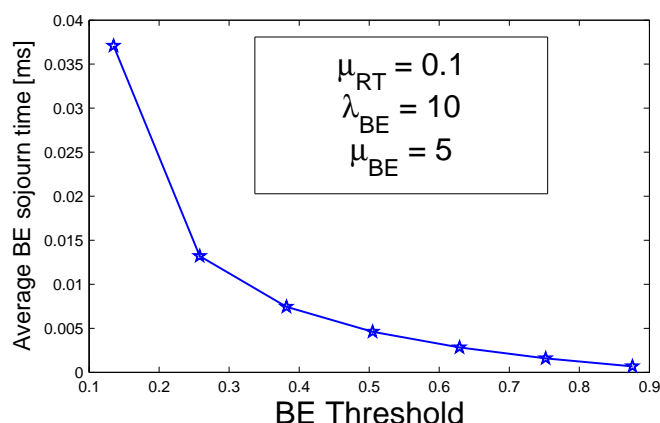


FIGURE 2.3 – Temps de séjour d'appels BE vs seuil de la capacité réservée aux appels BE, pour $\mu_{RT} = 0.1$, $\mu_{BE} = 5$, $\lambda_{BE} = 10$, et $\bar{\rho}_{RT} = 0.5$.

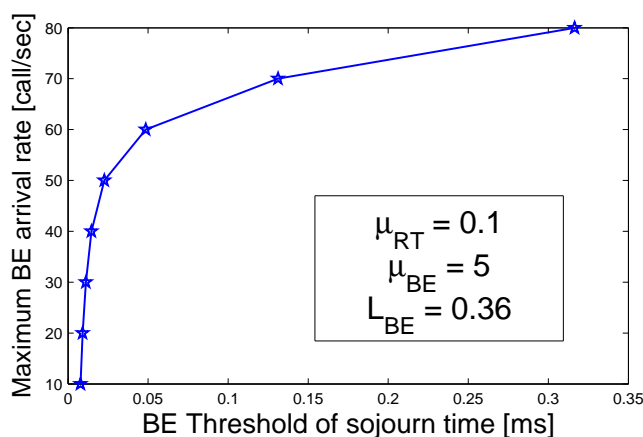


FIGURE 2.4 – Taux d'arrivée maximum d'appels BE en fonction du seuil de temps de séjour d'appels BE, pour $\mu_{RT} = 0.1$, $\mu_{BE} = 5$, et $\bar{\rho}_{RT} = 0.5$.

Validation des résultats par le simulateur NS2 : Dans ce paragraphe nous utilisons le simulateur à événements discrets (Network Simulator) NS2 pour valider nos résultats théoriques précédents. Nous avons considéré les paramètres suivants : la durée moyenne de service d'appels RT est de 180 s ; le taux d'arrivée d'appels RT est de 0.5

$appel/s$; la taille moyenne d'une session BE est de 1.6 Mbits et le taux d'arrivée d'appels BE est de 0.5 appel/s . Nous comparons les résultats analytiques et ceux issues de la simulation dans les figures 2.5 et 2.6. Par conséquent, nous remarquons dans ces figures que les deux résultats, analytiques et simulations sont quasiment les mêmes, ce qui valide notre approche théorique.

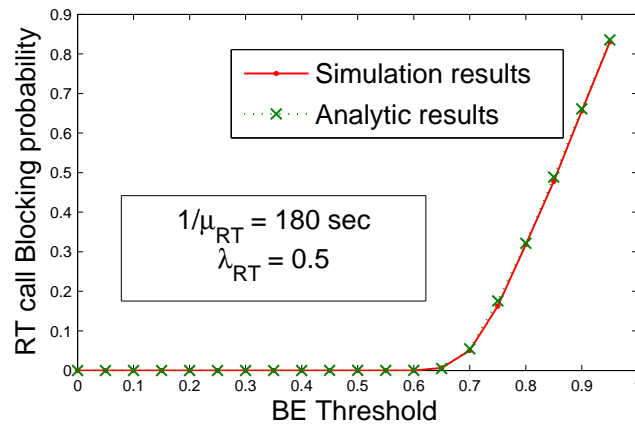


FIGURE 2.5 – Probabilité de blocage d'appels RT en fonction du seuil L_{BE} , comparaison entre les résultats analytiques et de simulations sur NS2.

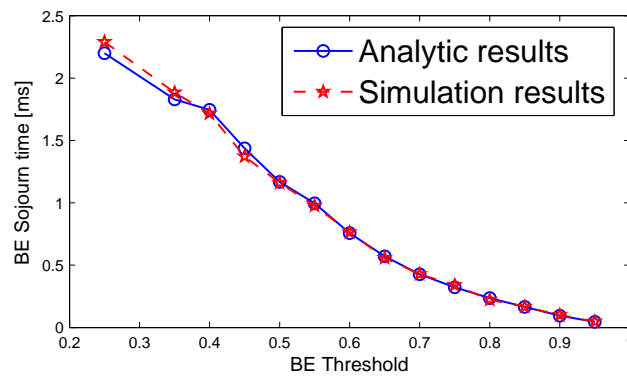


FIGURE 2.6 – Temps de séjour d'appels BE en fonction du seuil L_{BE} , comparaison entre les résultats analytiques et de simulations sur NS2.

2.3 Cas de trois classes de service

Puisque plusieurs trafics dans la classe BE ont besoin d'un minimum de qualité de service, la nécessité d'introduire un autre type de trafic est primordiale. En effet, une connexion FTP par exemple, a besoin d'un débit minimum pour échanger les fichiers entre le serveur et le client. Dans cette section, nous avons considéré plusieurs flots multimédia ayant chacun des besoins en qualité de service et des caractéristiques de trafic différents. Nous avons analysé les performances de la stratégie (contrôle d'admission) avec la priorité donnée à l'obtention d'une meilleure performance globale du système en respectant à peu près la qualité de service demandée par chaque classe. En effet, nous proposons trois classes de services pour mieux gérer les nouveaux services dans l'UMTS. Ces classes de qualité de services sont classées selon les applications :

- Trafic temps réel (RT) : pour des flux temps réel de paquets de taille fixe émis périodiquement, mode particulièrement adapté aux conversations vocales qui bénéficient d'une bande passante contrôlée avec échange interactif en temps réel et avec un minimum de délai entre les paquets ;
- Trafic non temps réel (NRT) pour des flux supposant des débits minimum mais tolérants en termes de délai comme les échanges en FTP ;
- Trafic *best effort* (BE) pour un trafic qui ne demande aucune qualité de service, mode particulièrement adapté au mode interactif qui est plutôt destiné à des échanges de type navigation sur le Web (requête/réponse).

Modèle : Nous modélisons le lien montant du système WCDMA multiservices par un processus de naissance et de mort d'une chaîne de Markov $(X(t), Y(t), Z(t))$ à trois dimensions. $X(t)$ correspond aux trafics RT, $Y(t)$ correspond aux trafics NRT et $Z(t)$ correspond aux trafics BE à l'instant t . Les taux d'arrivées de ces trafics sont respectivement λ_{RT} , λ_{NRT} et λ_{BE} . Ainsi les taux de départ de ces trafics sont respectivement μ_{RT} , μ_{NRT} et μ_{BE} . Nous supposons que les temps inter-arrivées, les durées d'appels RT, les tailles des fichiers NRT et BE sont mutuellement indépendants. Nous utilisons les mêmes notions utilisées dans le premier cas de deux classes 2.2. La capacité du système définie dans (2.8) devient ici

$$L_{RT} = \Theta_\epsilon - L_{NRT}, \quad (2.71)$$

où L_{NRT} est la capacité minimale nécessaire à réserver aux appels NRT pour avoir un débit minimum et L_{RT} les ressources utilisées par les RT.

2.3.1 Contrôle d'admission du trafic temps réel

Nous rappelons que l'UMTS utilise la technique AMR pour le trafic RT. AMR propose huit taux de transmission variant entre 4.75 kbps et 12.2 Kbps dynamiquement chaque 20 msec. Soient R_{min}^{RT} le taux minimum de transmission et R_{max}^{RT} le taux maximum de transmission. Donc les taux de transmission d'appels RT appartient à $[R_{min}^{RT}, R_{max}^{RT}]$. La bande minimale normalisée correspond au taux R_{min}^{RT} et est désignée par Δ_{min}^{RT} et celle

correspondant au R_{max}^{RT} est notée par Δ_{max}^{RT} . Leurs expressions sont

$$\begin{aligned}\Delta_{max}^{RT} &= \frac{\Gamma E_{RT}/N_0}{W/R_{max}^{RT} + \Gamma E_{RT}/N_0}, \\ \Delta_{min}^{RT} &= \frac{\Gamma E_{RT}/N_0}{W/R_{min}^{RT} + \Gamma E_{RT}/N_0},\end{aligned}$$

où E_{RT}/N_0 est l'énergie requise pour transmettre un bit de type RT. Le nombre d'appels RT correspondant au taux de transmission R_{min}^{RT} est donnée par $M_{RT} = \lfloor L_{RT}/\Delta_{max}^{RT} \rfloor$ et celui qui correspond au taux R_{max}^{RT} est $N_{RT} = \lfloor L_{RT}/\Delta_{min}^{RT} \rfloor$ où $\lfloor x \rfloor$ désigne la partie entière de x . Alors, la bande nécessaire pour un appel RT est calculée dynamiquement comme suit

$$\Delta(i) = \begin{cases} \Delta_{max}^{RT}, & \text{si } i \leq N_{RT}, \\ L_{RT}/i, & \text{si } N_{RT} < i \leq M_{RT}, \end{cases} \quad (2.72)$$

Puisque le processus $X(t)$ est indépendant des autres processus $Y(t)$ et $Z(t)$, nous le modélisons tout seul par un processus de naissance et de mort avec un taux de naissance λ_{RT} et un taux de mort μ_{RT} . Le débit d'appel RT est

$$R_{RT}(i) = \frac{\Delta(i)}{1 - \Delta(i)} \frac{W}{E_{RT}/N_0}.$$

Le débit moyen total d'appels temps réel est défini par

$$\bar{D}_{RT} = \frac{\sum_{i=0}^{M_{RT}} iP(X(t) = i)R_{RT}(i)}{E(X(t))},$$

où

$$E(X(t)) = \sum_{i=0}^{M_{RT}} iP(X(t) = i),$$

$$P(X(t) = i) = \frac{\rho_{RT}^{M_{RT}}/M_{RT}!}{\sum_{j=0}^i \frac{\rho_{RT}^j}{j!}},$$

où ρ_{RT} est le taux de charge d'appels RT défini par $\rho_{RT} = \frac{\lambda_{RT}}{\mu_{RT}}$. La probabilité de blocage d'un nouvel appel temps réel est

$$P_B^{RT} = \frac{\rho_{RT}^{M_{RT}}/M_{RT}!}{\sum_{i=0}^{M_{RT}} \frac{\rho_{RT}^i}{i!}}.$$

2.3.2 Contrôle d'admission du trafic non temps réel

Nous considérons que les appels NRT ont une capacité minimale non nulle L_{NRT} et peuvent augmenter leurs ressources dans le cas où il y a moins d'appels RT dans le système. Les ressources qui seront partagées par les appels NRT sont

$$L_{NRT}(i) = \begin{cases} \Theta_\epsilon - (N_{RT} - 1)\Delta_{max}^{RT}, & \text{si } i \leq N_{RT}, \\ L_{NRT}, & \text{si } N_{RT} < i \leq M_{RT}, \end{cases} \quad (2.73)$$

Nous définissons un contrôle d'admission pour les appels NRT, qui vise à garantir une qualité de service particulière. Pour avoir cette qualité de service, un nombre maximum M_{NRT} d'appels NRT devrait être accepté par le système. Ce nombre est défini comme un nombre qui correspond au taux d'arrivée maximal d'appels NRT tels que leurs temps de séjour prévu est borné par une constante positive imposée par l'opérateur.

Le taux de départ d'appels NRT dépend du nombre de ceux en RT dans le système. Alors, le taux effectif de départ d'appels NRT est donné par

$$v_1(i) = \mu_{NRT} R_{NRT}^{tot}(i),$$

où $R_{NRT}^{tot}(i)$ est le taux total de transmission d'appels NRT correspondant à leur capacité disponible $L_{NRT}(i)$ et est défini par

$$R_{NRT}^{tot}(i) = \frac{L_{NRT}(i)}{1 - L_{NRT}(i)} \times \frac{W}{\Gamma E_{NRT}/N_0},$$

où E_{NRT}/N_0 est l'énergie pour transmettre un bit de type NRT.

Modèle du trafic non temps réel : La chaîne de Markov $(X(t), Y(t))$ est un processus QBD de deux dimensions, dont les taux de naissances sont λ_{RT} et λ_{NRT} . Alors que ses taux de morts sont μ_{RT} et μ_{NRT} . Nous nous intéressons ici au calcul de la distribution stationnaire, notée \mathbf{y} de cette chaîne. Soit \mathbf{Q}_Y le générateur correspondant dont la dimension est $(M_{RT} + 1)(M_{NRT} + 1)$. La distribution est solution du système suivant

$$\mathbf{y}\mathbf{Q}_Y = \mathbf{0}, \mathbf{y}\mathbf{e} = 1, \quad (2.74)$$

où \mathbf{e} (resp, $\mathbf{0}$) est un vecteur colonne des uns (resp, zéros) de même dimension. Le vecteur \mathbf{y} représente les probabilité d'états d'équilibres du processus de naissance et de mort à deux dimensions et sa forme est $\{\mathbf{y}_0, \dots, \mathbf{y}_{M_{NRT}}\}$, où \mathbf{y}_j est un vecteur qui représente le niveau j . Chaque niveau peut partitionner en probabilité d'avoir i appels RT comme suit : $\mathbf{y}_j = [y_j(0), \dots, y_j(M_{RT})]$ et le générateur est de la forme

$$\mathbf{Q}_Y = \begin{bmatrix} \mathbf{B}_1 & \mathbf{A}_0 & 0 & \dots & \dots \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & 0 & \dots \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 \\ 0 & \dots & 0 & \mathbf{A}_2 & \mathbf{B}_2 \end{bmatrix},$$

où \mathbf{A}_0 , \mathbf{A}_1 , \mathbf{B}_1 , \mathbf{A}_2 et \mathbf{B}_2 sont des matrices carrées d'ordre $M_{RT} + 1$. La matrice $\mathbf{A}_0 = \text{diag}(\lambda_{NRT})$ représente les arrivées d'appels NRT. La matrice \mathbf{A}_2 quant à elle, représente les départs d'appels NRT, $\mathbf{A}_2 = \text{diag}(v_1(i), 0 \leq i \leq M_{RT})$. Nous définissons les matrices \mathbf{A}_1 , \mathbf{B}_2 et \mathbf{B}_1 sous la forme $\mathbf{A}_1 = \mathbf{C}_1 - \mathbf{A}_0 - \mathbf{A}_2$, $\mathbf{B}_2 = \mathbf{A}_1 + \mathbf{A}_0$ et $\mathbf{B}_1 = \mathbf{A}_1 + \mathbf{A}_2$. Ainsi la matrice \mathbf{C}_1 représente les arrivées et les départs d'appels RT dont la forme est

$$\begin{aligned} \mathbf{C}_1[i, i] &= -(\lambda_{RT} + i\mu_{RT}), 0 \leq i \leq M_{RT} - 1; \\ \mathbf{C}_1[i, i + 1] &= \lambda_{RT}, 0 \leq i \leq M_{RT} - 1; \\ \mathbf{C}_1[i, i - 1] &= i\mu_{RT}; 1 \leq i \leq M_{RT}; \\ \mathbf{C}_1[M_{RT}, M_{RT}] &= -M_{RT}\mu_{RT}. \end{aligned}$$

La distribution stationnaire de ce système est donnée par l'équation (2.61) dans le cas de deux classes sous la forme

$$\mathbf{y}_j = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^j w_i + \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i \left(\frac{1}{\phi_i}\right)^{M_{NRT}-j} w_i, \quad (2.75)$$

où $\phi_0, \phi_1, \phi_2, \dots, \phi_{M_{RT}}$ sont des valeurs propres correspondant aux vecteurs propres $w_0, \dots, w_{M_{RT}}$ de la matrice \mathbf{R} , tel que $w_i \mathbf{R} = \phi_i w_i$; Ainsi $\frac{1}{\phi_i}; i = M_{RT} + 1, \dots, 2M_{RT} + 2$ sont les valeurs propres associées aux vecteurs propres $w_{M_{RT}+1}, \dots, w_{2M_{RT}+2}$ de la matrice \mathbf{S} . Notons que ces matrices \mathbf{R} et \mathbf{S} sont données dans le cas de deux classes dans ce chapitre. Les inconnus β_i représentent les coefficients des vecteurs v_1 et v_2 et sont solutions du système d'équation suivant

$$\begin{aligned} \mathbf{y}_0 \mathbf{B}_1 + \mathbf{y}_1 \mathbf{A}_2 &= \mathbf{0}, \\ \mathbf{y}_{M_{NRT}-1} \mathbf{A}_0 + \mathbf{y}_{M_{NRT}} \mathbf{B}_2 &= \mathbf{0}, \end{aligned}$$

et

$$\mathbf{y} \mathbf{e} = 1.$$

Pour résoudre le système ci-dessus, nous pouvons l'écrire sous forme matricielle

$$(\beta_0, \dots, \beta_{2M_{RT}+2}) \times \begin{bmatrix} \mathbf{M1} & \mathbf{M2} \\ \mathbf{M3} & \mathbf{M4} \end{bmatrix} = (\mathbf{0}, \mathbf{0}), \quad (2.76)$$

où

$$\begin{aligned} \mathbf{M1} &= \mathbf{V} \mathbf{B}_1 + \mathbf{\Psi} \mathbf{V} \mathbf{A}_2, \\ \mathbf{M2} &= \mathbf{\Psi}^{(M_{NRT}-1)} \mathbf{V} \mathbf{A}_0 + \mathbf{\Psi}^{M_{NRT}} \mathbf{V} \mathbf{B}_2, \\ \mathbf{M3} &= \mathbf{\Phi}^{M_{NRT}} \mathbf{W} \mathbf{B}_1 + \mathbf{\Phi}^{(M_{NRT}-1)} \mathbf{W} \mathbf{A}_2, \\ \mathbf{M4} &= \mathbf{\Phi} \mathbf{W} \mathbf{A}_0 + \mathbf{W} \mathbf{B}_2, \\ \mathbf{\Psi} &= \text{diag}(\phi_0, \phi_1, \dots, \phi_{M_{RT}}), \\ \mathbf{\Phi} &= \text{diag}(1/\phi_{M_{RT}+1}, \dots, 1/\phi_{2M_{RT}+2}), \\ \mathbf{V} &= (w_0, \dots, w_{M_{RT}}), \\ \mathbf{W} &= (w_{M_{RT}+1}, \dots, w_{2M_{RT}+2}). \end{aligned}$$

Ainsi la condition de normalisation devient

$$\sum_{i=0}^x \beta_i \frac{1 - \phi_i^{(M_{NRT}+1)}}{1 - \phi_i} w_i \bar{\mathbf{1}} + \sum_{i=x+1}^{2x+2} \beta_i \frac{1 - (1/\phi_i)^{(M_{NRT}+1)}}{1 - 1/\phi_i} w_i \bar{\mathbf{1}} = 1, \quad (2.77)$$

où $x = M_{RT}$ et $\bar{\mathbf{1}}$ est un vecteur colonne des uns de même dimension que le vecteur w_i . Pour $z = 1$, nous utilisons la relation $\frac{1 - z^{y+1}}{1 - z} = y + 1$, car on a $1 - z^{y+1} = (1 - z)(1 + z + \dots + z^y)$. Par conséquent, nous combinons les équations (2.76) et (2.77) et trouvons les coefficients β_i comme solutions du système suivant

$$(\beta_0, \dots, \beta_{2x+2}) \times \begin{bmatrix} \mathbf{Vr} & \mathbf{M1} & \mathbf{M2} \\ \mathbf{Vs} & \mathbf{M3} & \mathbf{M4} \end{bmatrix} = (1, \mathbf{0}, \mathbf{0}),$$

où

$$\mathbf{Vr} = \left(\frac{1 - \phi_0^{(M_{NRT}+1)}}{1 - \phi_0} w_0 \bar{\mathbf{1}}, \dots, \frac{1 - \phi_x^{(M_{NRT}+1)}}{1 - \phi_x} w_x \bar{\mathbf{1}} \right)^T,$$

et

$$\mathbf{Vs} = \left(\frac{1 - \left(\frac{1}{\phi_{x+1}}\right)^{(M_{NRT}+1)}}{1 - \frac{1}{\phi_{x+1}}} w_{x+1} \bar{\mathbf{1}}, \dots, \frac{1 - \left(\frac{1}{\phi_{2x+2}}\right)^{(M_{NRT}+1)}}{1 - \frac{1}{\phi_{2x+2}}} w_{2x+2} \bar{\mathbf{1}} \right)^T.$$

Le nombre moyen d'appels NRT est calculé par $E[Y] = \sum_{j=0}^{M_{NRT}} j y_j e$ et la probabilité de blocage d'un nouvel appel NRT est déterminée quand le nombre d'appels dans le système est exactement M_{NRT} via l'équation (2.75) :

$$P_B^{NRT} = Pr[Y = M_{NRT}] = \sum_{i=0}^{M_{RT}} \beta_i (\phi_i)^{M_{NRT}} w_i e + \mathbf{v}_2 e, \quad (2.78)$$

où $\mathbf{v}_2 = \sum_{i=M_{RT}+1}^{2M_{RT}+2} \beta_i w_i$ et le temps de séjour d'appels NRT est obtenu via la formule de Little

$$T^{NRT} = \frac{E[Y]}{\lambda_{NRT}(1 - P_B^{NRT})}. \quad (2.79)$$

Nous rappelons que le délai T^{NRT} d'appels NRT est inférieure au égale à une constante imposée par l'opérateur qu'on note par d . Cela signifie que $T^{NRT}(M_{NRT}^*) \leq d$. Alors le nombre maximum d'appels NRT que peut manipuler le système en satisfaisant cette qualité de service est noté M_{NRT}^* . Enfin, ce nombre M_{NRT}^* est exactement M_{NRT} .

2.3.3 Trafic *best effort* sans contrôle d'admission

Nous rappelons que les appels BE dépendent de ceux en RT et en NRT. Nous supposons que les appels BE n'exigent aucune qualité de service et adaptent leurs taux

de transmission selon les ressources du réseau. Ainsi, ils partagent les ressources non utilisées par les appels NRT, d'où le temps de départ d'appels BE défini par

$$v_2(i, j) = \mu_{BE} R_{BE}^{tot}(i, j),$$

où $R_{BE}^{tot}(i, j)$ est le taux total de transmission correspondant à la capacité disponible notée $C(i, j)$, dont sa forme est

$$R_{BE}^{tot}(i, j) = \frac{C(i, j)}{1 - C(i, j)} \times \frac{W}{\Gamma E_{BE}/N_0}.$$

où E_{BE}/N_0 est l'énergie pour transmettre un bit de type BE sur la densité du bruit et

$$C(i, j) = \begin{cases} L_{NRT}(i) - j\Delta_{NRT}(i), & \text{si } j \leq M_{NRT}, \\ 0, & \text{si } N_{RT} < i \leq M_{RT}, \end{cases} \quad (2.80)$$

où $\Delta_{NRT}(i)$ ($i = 0, 1, \dots, M_{RT}$) est la capacité requise par appel NRT $\Delta_{NRT}(i) = \frac{L_{NRT}(i)}{M_{NRT}}$. Puisque nous avons supposé que les appels BE n'ont aucun contrôle d'admission, le système accepte un nombre infini parmi eux. La chaîne de Markov $(X(t), Y(t), Z(t))$ dont le générateur noté \mathbf{Q} est défini par

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P}_0 & \mathbf{Q}_0 & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{Q}_2 & \mathbf{Q}_1 & \mathbf{Q}_0 & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{Q}_2 & \mathbf{Q}_1 & \mathbf{Q}_0 & \dots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \ddots \end{bmatrix}$$

Les matrices $\mathbf{P}_0, \mathbf{Q}_0, \mathbf{Q}_1$ et \mathbf{Q}_2 sont des matrices carrées par bloc d'ordre $M_{NRT} + 1$. Leurs formes sont

$$\mathbf{Q}_0 = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{A}_{11} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{A}_{11} \end{bmatrix}$$

$$\mathbf{Q}_2 = \begin{bmatrix} \mathbf{A}_2^{(0)} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{A}_2^{(1)} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{A}_2^{(M_{NRT})} \end{bmatrix}$$

$$\mathbf{Q}_1 = \mathbf{Q}_Y - \mathbf{Q}_2 - \mathbf{Q}_0, \quad \mathbf{P}_0 = \mathbf{Q}_Y - \mathbf{Q}_0$$

où \mathbf{A}_{11} est une matrice carrée d'ordre $M_{RT} + 1$ de la forme $\mathbf{A}_{11} = \text{diag}(\lambda_{BE})$. La matrice \mathbf{Q}_2 contient les taux de départs d'appels BE et formée des blocs matricielle carrées $\mathbf{A}_2^{(j)}$ d'ordre $M_{RT} + 1$ de la forme $\mathbf{A}_2^{(j)} = \text{diag}(v_2(i, j) \quad 0 \leq i \leq M_{RT}); 0 \leq j \leq M_{NRT}$. Enfin, les matrices \mathbf{P}_0 et \mathbf{Q}_1 sont tridiagonales, alors que \mathbf{Q}_0 et \mathbf{Q}_2 sont des matrices diagonales.

2.3.4 Évaluation des performances par l'analyse spectrale

Nous appliquons l'analyse spectrale dans le cas d'une chaîne de Markov à trois dimensions. Cette approche va nous permettre de trouver explicitement de la distribution stationnaire de cette chaîne. Cette distribution de probabilités, notée par π , est la solution du système suivant

$$\pi \mathbf{Q} = \mathbf{0}, \quad (2.81)$$

$$\pi \mathbf{e} = 1, \quad (2.82)$$

où \mathbf{e} est un vecteur d'ordre infini des uns, $\pi = [\pi^0, \pi^1, \dots, \pi^{n-1}, \pi^n, \dots]$, π^k est un vecteur dont les composantes sont données par la probabilité d'avoir i RT, j NRT et k BE appels ($\pi_{i,j,k} = P(X = i, Y = j, Z = k)$). Nous avons également

$$\pi^0 \mathbf{P}_0 + \pi^1 \mathbf{Q}_2 = \mathbf{0}, \quad (2.83)$$

$$\pi^{n-1} \mathbf{Q}_0 + \pi^n \mathbf{Q}_1 + \pi^{n+1} \mathbf{Q}_2 = \mathbf{0}, \quad n \geq 1. \quad (2.84)$$

Nous décomposons le vecteur de probabilités de la ligne n comme suit $\pi^n = [\pi_{0,n}, \dots, \pi_{M_{NRT},n}]$. Chaque composante de ce dernier peut s'écrire sous la forme $\pi_{j,n} = [\pi_{0,j,n}, \dots, \pi_{M_{RT},j,n}]$ pour j NRT et n BE appels.

Théorème 2 *Le système est stable si et seulement si*

$$\lambda_{BE} < \sum_{j=0}^{M_{NRT}} \sum_{i=0}^{M_{RT}} x_i^j v_2(i, j) = c, \quad (2.85)$$

où $x_i^j = \pi_{i,j,n}$.

Démonstration du Théorème 2 *Pour avoir la démonstration, nous appliquons le théorème 1. 7. 1 qui se trouve dans (Neuts, 1981). Donc, la condition de stabilité du système est définie comme le taux moyen d'arrivée qui doit être inférieur au taux moyen de service du système, ce qui signifie que $\mathbf{y} \mathbf{Q}_0 \mathbf{e}_y < \mathbf{y} \mathbf{Q}_2 \mathbf{e}_y$, où \mathbf{y} est la distribution stationnaire du système (2.74), ainsi nous avons $\mathbf{Q}_y \mathbf{e} = \mathbf{0}$, car la matrice \mathbf{Q}_y est stochastique. Par conséquent, le générateur \mathbf{Q} est une matrice stochastique et le système correspondant est stable si et seulement si*

$$\lambda_{BE} < \sum_{j=0}^{M_{NRT}} \sum_{i=0}^{M_{RT}} x_i^j v_2(i, j),$$

où $x_i^j = \pi_{i,j,n}$.

Nous déduisons que la chaîne de Markov décrite par le générateur ci-dessus, est récurrente positive sous la condition de stabilité (2.85). D'où, la solution (2.81) est unique et satisfait la condition de normalisation (2.82). Nous supposons que le système est stable. Ecrivons sa solution sous la forme développée par Latouche (Latouche et Ramaswami, 1999) : $\pi^n = \pi^0 \mathbf{R}^n$, où \mathbf{R} est une matrice géométrique qui désigne la solution minimale non négative de l'équation matricielle quadratique suivante

$$\mathbf{R}^2 \mathbf{Q}_2 + \mathbf{R} \mathbf{Q}_1 + \mathbf{Q}_0 = \mathbf{0}. \quad (2.86)$$

et le vecteur de probabilité π^0 est obtenu en résolvant le système

$$\pi^0(\mathbf{I} - \mathbf{R})^{-1} = \mathbf{y} = \sum_{n=0}^{\infty} \pi^n,$$

où \mathbf{y} est un vecteur de probabilité donné dans (2.75). Les valeurs propres de la matrice \mathbf{R} sont les racines de l'équation

$$\det(\mathbf{T}(z)) = 0,$$

$$\mathbf{T}(z) = \mathbf{Q}_0 + z\mathbf{Q}_1 + z^2\mathbf{Q}_2. \quad (2.87)$$

$\mathbf{T}(z)$ est donc une matrice polynômiale tridiagonale composée de $M_{NRT} + 1$ blocs en lignes et en colonnes, dont la taille est $N = (M_{NRT} + 1)(M_{RT} + 1)$ et les composants sont donnés par

$$\begin{aligned} \mathbf{T}(z)[j, j] &= \mathbf{A}_{11} + z(\mathbf{A}_1 - \mathbf{A}_0 - \mathbf{A}_2) + z^2\mathbf{A}_2^{(j)}, \quad \forall j; \\ \mathbf{T}(z)[j, j-1] &= z\mathbf{A}_2, \quad j \in (1, \dots, M_{NRT}); \\ \mathbf{T}(z)[j, j+1] &= z\mathbf{A}_0, \quad j \in (0, \dots, M_{NRT} - 1). \end{aligned}$$

Nous réécrivons la matrice $\mathbf{T}(z)$ sous la nouvelle forme

$$\mathbf{T}(z) = (z\mathbf{I} - \mathbf{R})(z\mathbf{Q}_2 + \mathbf{R}\mathbf{Q}_2 + \mathbf{Q}_1). \quad (2.88)$$

L'équation suivante (2.89) est très intéressante puisqu'elle montre les valeurs propres de la matrice \mathbf{R} comme racines du polynôme $\det[\mathbf{T}(z)]$.

$$\det[\mathbf{T}(z)] = P_{\mathbf{R}}(z) \times f(z), \quad (2.89)$$

où $f(z) = \det(z\mathbf{Q}_2 + \mathbf{R}\mathbf{Q}_2 + \mathbf{Q}_1)$ et $P_{\mathbf{R}}(z) = \det(z\mathbf{I} - \mathbf{R})$. Sous la condition de stabilité (2.85), le rayon spectrale de \mathbf{R} est inférieur à 1.

- Lemme 4**
1. Si la valeur λ est une valeur propre de la matrice \mathbf{R} , dont la multiplicité est n . Alors elle est également racine de multiplicité au moins égale à n du polynôme $\det[\mathbf{T}(z)]$.
 2. Si le vecteur v est un vecteur propre associé à la valeur propre λ , alors v est un vecteur qu'annule $\mathbf{T}(\lambda)$ à gauche, cela signifie que $v\mathbf{T}(\lambda) = 0$.
 3. Les matrices carrées \mathbf{R} et \mathbf{Q}_0 ont même noyau à gauche.

Démonstration du Lemme 4 Les démonstartions sont similaires à celles des lemmes du cas de deux classes 2.2.4 de ce chapitre.

Lemme 5 Toutes les valeurs propres de la matrice \mathbf{R} sont dans le disque de rayon 1.

Démonstration du Lemme 5 Puisque la matrice \mathbf{Q}_Y est irréductible et en appliquant les théorèmes 1.3.1 et 3.1.1 cités dans (Neuts, 1981), nous trouvons $sp(\mathbf{R}) < 1$.

Théorème 3 Le polynôme $\det[\mathbf{T}(z)]$ a au plus $2N$ racines positives et pour leur localisation, nous distinguons les deux cas :

1. Si $\lambda_{BE} < c$, alors toutes les valeurs propres sont simples et non nulles. N d'entre eux se trouvent dans le disque de rayon 1, une racine à $z = 1$ et N sont localisées à l'extérieur du disque unité.

2. Si $\lambda_{BE} = c$, alors toutes les valeurs propres sont simples et non nulles. Deux à $z = 1$, $N - 1$ dans le disque unité et $N - 1$ à l'extérieur du disque unité.

Démonstration du Théorème 3 1. Nous remarquons que le processus QBD du générateur \mathbf{Q} est récurrent positif (Neuts, 1981), car $\lambda_{BE} < c$. Soit \mathbf{G} la solution minimale non négative de l'équation quadratique

$$\mathbf{Q}_0 z^2 + \mathbf{Q}_1 z + \mathbf{Q}_2 = \mathbf{0}.$$

Et $\mathbf{F}(z)$ une matrice polynomiale de la forme

$$\mathbf{F}(z) = z^2 \mathbf{Q}_0 + z \mathbf{Q}_1 + \mathbf{Q}_2 = (z \mathbf{Q}_0 + \mathbf{Q}_1 + \mathbf{Q}_0 \mathbf{G})(z \mathbf{I} - \mathbf{G}).$$

Nous écrivons le polynôme $\mathbf{T}(z)$ en fonction de \mathbf{F} comme suit

$$\mathbf{T}(z) = z^2 \mathbf{F}(1/z).$$

Alors nous obtenons la relation suivante (voir (Lancaster, 1984))

$$\det[\mathbf{T}(z)] = z^{2N} \det[\mathbf{F}(z^{-1})].$$

Ce qui montre que le polynôme $\det[\mathbf{T}(z)]$ a au moins un degré d'ordre $2N$. La matrice $\mathbf{T}(z)$ est réécrite sous la forme

$$\mathbf{T}(z) = (1 - z)(\mathbf{Q}_0 - z \mathbf{Q}_2) + z \mathbf{Q}_Y.$$

Nous trouvons quand $z = 1$ l'égalité $\det[\mathbf{T}(1)] = \det(\mathbf{Q}_Y) = 0$, car la matrice \mathbf{Q}_Y définie dans (2.74) est une matrice stochastique singulière, c'est à dire qu'il existe un vecteur non nul dans notre cas est bien le vecteur de probabilité \mathbf{y} tel que $\mathbf{y} \mathbf{Q}_Y = \mathbf{0}$. D'où le polynôme $\det[\mathbf{T}(z)]$ a une racine à $z = 1$. En plus, selon l'équation (2.88) les valeurs propres de \mathbf{R} sont exactement les racines du polynôme $\det[\mathbf{T}(z)]$ et sont toutes inférieurs à 1 via le Lemme 5. Donc, le polynôme $\det[\mathbf{T}(z)]$ a N racines quand $|z| < 1$ et $N - 1$ quand $|z| > 1$.

2. Nous utilisons la même méthode que celle utilisée avant. En effet, si $\lambda_{BE} = c$ le processus QBD est récurrent nul comme montré dans (Neuts, 1981). D'où nous avons $sp(\mathbf{R}) = 1$, ce qui montre qu'il y a une autre racine à $z = 1$ du polynôme $\det[\mathbf{T}(z)]$.

Après avoir présenté l'analyse du système par l'approche d'analyse spectrale, nous écrivons la distribution stationnaire π^n sous la forme spectrale suivante

$$\pi^n = \sum_{i=1}^N \alpha_i (\psi_i)^n v_i, \quad n = 0, 1, 2, 3, \dots \quad (2.90)$$

où $N = (M_{RT} + 1) \times (M_{NRT} + 1)$; ψ_0, \dots, ψ_N sont les valeurs propres de la matrice \mathbf{R} associées aux vecteurs propres v_0, \dots, v_N , liées par les relations $v_i \mathbf{R} = \psi_i v_i$ ($i = 0, \dots, N$) et $\alpha_0, \dots, \alpha_N$ sont des coefficients réels. Nous calculons les valeurs propres ψ_i en utilisant le Lemme 4 comme racines du polynôme $\det[\mathbf{T}(z)]$ qui se trouvent dans le disque unité pour $|z| < 1$. Aussi, les coefficients α_i sont obtenus en utilisant la condition de normalisation (2.82).

Nous présentons quelques métriques du système. Soit z_n la probabilité marginale d'appels BE défini par

$$z_n = \pi^n e,$$

où e est un vecteur colonne des uns de même dimensions que les vecteurs propres. Ainsi, le nombre moyen marginal d'appels BE est calculé par

$$E[Z] = \sum_{n=0}^{\infty} n z_n.$$

Ensuite, le temps moyen de séjour d'appels BE est trouvé en appliquant le résultat de Little $T^{BE} = \frac{E[Z]}{\lambda_{BE}}$. Sa forme en fonction des valeurs et vecteurs propres de la matrice R est

$$T^{BE} = \frac{1}{\lambda_{BE}} \sum_{i=1}^N \alpha_i \frac{\psi_i}{(1 - \psi_i)^2} v_i e. \quad (2.91)$$

2.3.5 Illustrations numériques

Nous proposons les évaluation des performances du système sous forme d'applications numériques en nous focalisant sur les probabilités de blocage et les temps de séjour. Nous utilisons les valeurs numériques suivantes (Holma et Toskala, 2001) : le débit minimum d'appels RT est 4.75 *kbps* et la bande nécessaire est $\Delta_{min}^{RT} = 0.124$; le débit maximum d'appels RT est 12.2 *kbps* et la bande correspondante est $\Delta_{max}^{RT} = 0.141$; les énergies transmises d'un bit de type RT (respectivement NRT et BE) sur la densité du bruit sont $E_{RT}/N_0 = 4.1$ dB (respectivement $E_{NRT}/N_0 = 4$ dB et $E_{BE}/N_0 = 3.1$ dB) ; le débit chip est $W = 3.84$ MHz et $\epsilon = 10^{-5}$ (Hegde et Altman, 2003). Le taux total de transmission d'appels RT (R_T) est défini quand tous ces appels ont servi avec un taux minimum R_{RT}^{min} par $R_T = \left| \frac{\Theta \epsilon}{\Delta_{min}^{RT}} \right| R_{min}^{RT}$. Nous considérons que le taux de charge normalisé des trafics temps réel ($\tilde{\rho}_{RT} = \rho_{RT} \frac{R_{max}^{RT}}{R_T}$) est relativement grand $\tilde{\rho}_{RT} = 0.3$. Nous rappelons que pour garantir un taux de transmission minimal non nul aux appels NRT, nous allons leur réserver une proportion minimale L_{NRT} de la capacité totale. Nous supposons que l'opérateur peut fixer cette proportion selon sa politique de gestion de ressources entre les trois classes de service considérées. Par exemple, il veut que cette proportion soit entre 10% et 90% de la capacité totale. Dans ce qui suit nous essayons de trouver un compromis entre tous les appels via les métriques de performances étudiées.

Dans ce paragraphe, nous nous basons sur les détails des formules illustrées dans les précédentes sous sections 2.3.1 et 2.3.2. La figure 2.7 montre la probabilité de blocage P_B^{RT} d'appels RT en fonction de la capacité minimale ou seuil L_{NRT} d'appels NRT. Nous remarquons que lorsque la capacité minimale augmente la probabilité de blocage croît aussi comme prévu. Cela se justifie par le fait qu'augmenter la capacité minimale pour les appels NRT diminue les ressources restantes de ceux en RT. Nous constatons dans cette figure que la probabilité de blocage reste constante et prend la valeur 15% quand

le seuil L_{NRT} appartenant à la région [40%, 50%]. En contre partie, quand L_{NRT} devient supérieur à 0.5, la probabilité de blocage d'appels RT progresse rapidement.

Ensuite, nous étudions l'effet de la capacité minimale L_{NRT} accordée aux appels NRT sur le temps de leur séjour et leur blocage. La figure 2.8 expose cet effet sur le temps de séjour d'appels NRT. Pour des petites valeurs de L_{NRT} (moins de 0.4), les appels NRT restent longtemps dans le système et pour les grandes valeurs de L_{NRT} (supérieur à 0.4), ces appels prennent moins de temps dans leurs service. Le résultat de cette figure montre que pour obtenir une très bonne configuration pour les appels NRT il faut leur réserver une bande passante importante. La figure 2.9 quant à elle, représente l'effet de variation des ressources minimales réservées aux appels NRT sur leur probabilité de blocage dans une échelle logarithmique. Dans la région $0 \leq L_{NRT} \leq 0.5$ la probabilité de blocage décroît lentement avec L_{NRT} . En dehors de cette région, la probabilité décroît rapidement en raison de l'augmentation des ressources disponibles aux appels NRT.

Donc, également nous constatons que la région [40%, 50%] donne de très bonnes performances pour les deux classes RT et NRT, moins de blocages et un temps de séjour acceptable. Il reste à voir les performances de la troisième classe BE dans cette région. Pour cela nous illustrons dans la figure 2.10 l'effet de la capacité minimale sur le temps de séjour d'appels BE. Comme nous avons prévu durant notre analyse théorique, le temps de séjour décroît quand les ressources réservées aux appels NRT augmentent. Nous remarquons aussi que la région trouvé avant est aussi importante pour la classe BE. Par conséquence, cette région offre un bon compromis entre les trois classes, nous avons moins de blocage pour RT et NRT et des temps de séjour raisonnables pour NRT et BE. Un autre résultat important est de proposer un CAC dans cette région. Dans la pratique, un fournisseur de services de télécoms ou opérateur peut appliquer un CAC sur les NRT et BE en fixant la capacité minimale L_{NRT} dans la région [40%, 50%]. Nous fixons $L_{NRT} = 0.4$ et proposons ici un CAC pour NRT et BE qui accepte un taux maximum d'arrivée NRT et BE en leurs garantissant une qualité de service particulière en terme de temps séjour fixé par l'opérateur comme le montrent les figures 2.11 et 2.12.

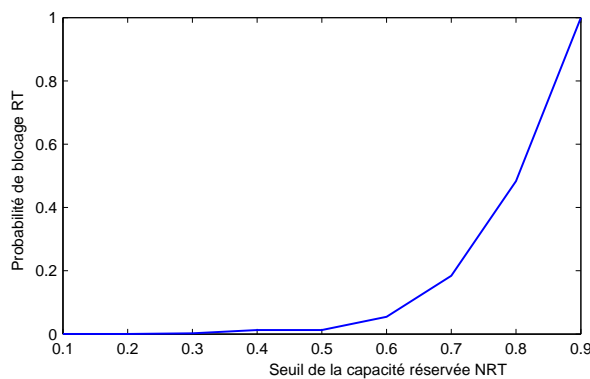


FIGURE 2.7 – Probabilité de blocage d'appels RT en fonction du seuil de la capacité réservée aux appels BE pour $\tilde{\rho}_{RT} = 0.3$ et $\mu_{RT} = 0.01$.

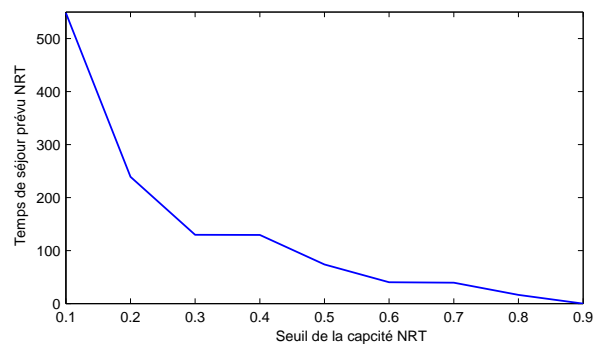


FIGURE 2.8 – Temps de séjour prévu d'appels NRT RT en fonction du seuil NRT pour $\tilde{\rho}_{RT} = 0.3$, $\mu_{RT} = 1$, $\mu_{NRT} = 0.0003$ et $\lambda_{NRT} = 5$.

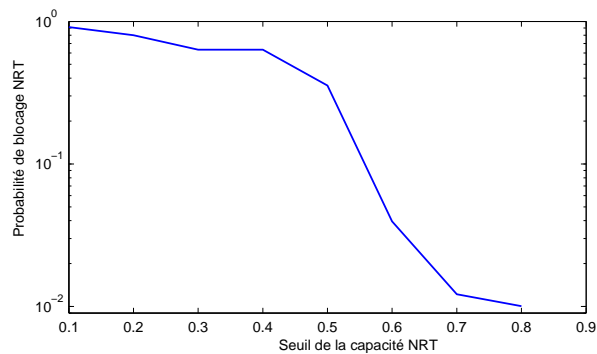


FIGURE 2.9 – Blocage d'appels NRT dans une échelle logarithmique RT en fonction du seuil NRT pour $M_{NRT} = 100$, $\lambda_{NRT} = 8$, $\mu_{RT} = 0.1$, $\tilde{\rho}_{RT} = 0.3$ et $\mu_{NRT} = 0.1$.

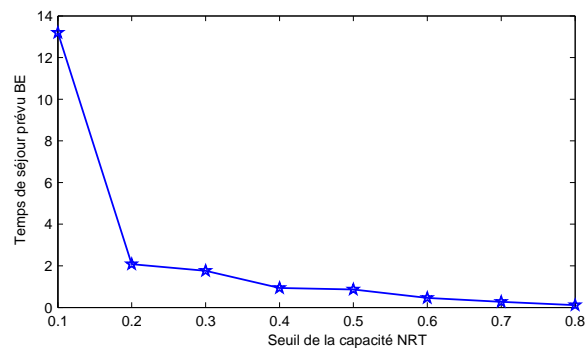


FIGURE 2.10 – Temps de séjour prévu pour les appels BE RT en fonction du seuil NRT pour $\lambda_{NRT} = 0.5$, $\mu_{NRT} = 0.1$, $\mu_{RT} = 0.1$, $\tilde{\rho}_{RT} = 0.3$, $\lambda_{BE} = 0.4$ et $\mu_{BE} = 0.1$.

2.4 Conclusion

Dans ce chapitre, nous avons effectué une analyse de l'évaluation de la capacité pour le lien montant du système WCDMA. Cette analyse est basée sur l'approche d'analyse

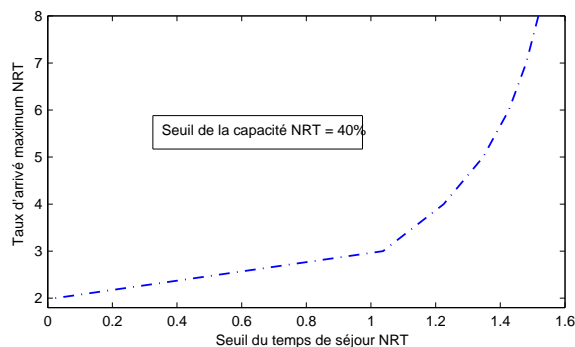


FIGURE 2.11 – Taux maximum d'arrivée d'appels NRT RT en fonction du seuil du temps de séjour NRT pour $\tilde{\rho}_{RT} = 0.3$, $\mu_{RT} = 0.1$ et $\mu_{NRT} = 0.1$.

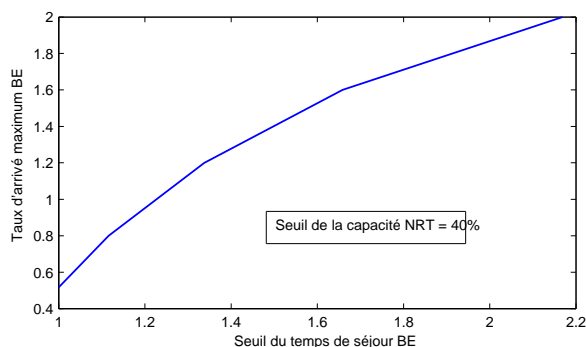


FIGURE 2.12 – Taux maximum d'arrivée d'appels BE RT en fonction du seuil du temps de séjour de ces appels pour $\mu_{RT} = 0.1$, $\tilde{\rho}_{RT} = 0.3$, $\lambda_{NRT} = 0.5$, $\mu_{NRT} = 0.1$ et $\mu_{BE} = 0.1$.

spectrale en séparant les services en deux classes (temps réel et *best effort*) et en trois classes (temps réel, non temps réel et *best effort*). Le cas de deux classes étudiées dans la première partie de ce chapitre nous a permis d'améliorer la gestion de la qualité de service en réduisant la complexité d'analyse des différentes classes de trafic. Car, la complexité du système du réseau de la troisième génération réside dans la diversité de services qu'il offre. L'analyse spectrale nous a permis de trouver de manière explicite la distribution des états d'équilibre du système dans deux scénarios du nombre d'appels *best effort* (fini et infini) que nous avons considéré. Nous signalons que la solution de la distribution stationnaire des états d'équilibre trouvée dans le cas de deux classes est explicite par rapport à celle proposée dans (Hegde et Altman, 2003). En outre, nous avons étudié l'influence de la bande passante minimale réservée aux appels BE sur les performances du système. De plus, nous avons fourni quelques résultats numériques de la performance du système et validé ces résultats par le biais d'un simulateur NS2.

Dans la deuxième partie de ce chapitre, nous avons modélisé le lien montant du système WCDMA, en considérant les trois classes de services temps réel, non temps réel et *best effort* pour réduire la complexité de gestion de ces services. Étant donné que certains services ont besoin d'un minimum de taux transmission à garantir par le fournisseur de service, nous avons introduit une classe de service intermédiaire entre les

deux classes étudiées dans la littérature temps réel et *best effort*. Cette classe, appelée non temps réel (NRT), regroupe tous les services qui demandent un débit minimum comme le transfert de fichiers (FTP). En utilisant également, l'approche d'analyse spectrale, nous avons analysé l'effet de la répartition de la capacité réservée aux appels non temps réel sur les performances des trois classes, nous avons obtenu des expressions explicites des performances et proposé un contrôle d'admission pour les appels non temps réel. Aussi, une analyse numérique est développée dans cette section et nous avons trouvé une région qui représente un compromis entre les trois classes étudiées

Chapitre 3

Évaluation des performances par la Perturbation Singulière

Sommaire

3.1 Introduction	79
3.2 Formulation du problème	81
3.2.1 Hypothèses	81
3.3 Série de Taylor	83
3.3.1 Définitions et Hypothèses	83
3.3.2 Représentation analytique de la distribution stationnaire	86
3.3.4 Temps de séjour d'appels <i>best effort</i>	88
3.4 Résultats numériques	89
3.5 Conclusion	93

3.1 Introduction

Dans le chapitre précédent, nous avons proposé une analyse très puissante et exacte de la capacité du réseau de troisième génération UMTS en s'appuyant sur la technique d'accès multiple WCDMA. Certes, cette analyse intitulée Analyse spectrale, nous a permis d'évaluer les performances du système qui jouent un rôle très important dans la garantie de la qualité de service demandée par les utilisateurs du réseau. Néanmoins, l'analyse spectrale demande un temps de calcul considérable vu la complexité de ses équations mathématiques quand le nombre d'appels de la classe temps réel (RT) devient grand. Cette complexité devient compliquée avec la diversité de service qu'offre le système aux utilisateurs mobiles.

Nous observons dans la plupart des cas, que les transitions du nombre d'appels RT se produisent beaucoup plus rapidement que les transitions du nombre d'appels BE. Dans ce cas particulier des comportements de transitions, nous définissons deux échelles de

transitions différentes, une rapide (nombre d'appels RT qui se produisent rapidement) et une lente (nombre d'appels BE qui se produisent lentement). Ce qui nous a permis de proposer une approximation simple et efficace en temps de calcul dans les conditions de la distribution stationnaire du système.

En effet, cette approximation appelée perturbation singulière est traitée dans ce chapitre. Nous considérons le système WCDMA en présence de deux classes de service : temps réel (RT) et *best effort* (BE). Les services de la classe RT ont des ressources dédiées et donc un contrôle d'admission, et les services de la classe BE n'ont pas de contrôle d'admission. Le premier objectif de ce chapitre est de calculer la distribution stationnaire de manière approximative des états d'équilibre du système en nombre d'appels RT et BE. Après cela, nous fournissons explicitement les performances du système. Néanmoins, le fournisseur de service ou l'opérateur peut utiliser cette approche basée sur l'approximation pour évaluer en temps réel avec un calcul rapide des métriques de performances d'appels RT et BE.

La perturbation singulière a été définie dans la littérature comme une faible interaction entre les groupes sachant que la forte interaction entre les états de chaque groupe est rapide. La perturbation singulière joue un rôle important dans les chaînes de Markov et surtout pour pouvoir déterminer une approximation simple des probabilités d'états d'équilibre dans lesquels nous pouvons identifier deux échelles de temps : une rapide et une lente. Les transitions entre les groupes sont beaucoup moins fréquentes que celles entre les états dans un même groupe ou bien l'inverse. La perturbation singulière a fait l'objet de plusieurs travaux de recherche et nous citons par exemple les travaux (Avrachenkov, 2004, 1999; Philips et al., 1981; Altman et al., 2004b; Avrachenkov et Haviv, 2004). Les auteurs dans (Philips et al., 1981), montrent que les chaînes de Markov finies admettant des interactions faibles et fortes, peuvent être décomposées en deux échelles de temps, rapide et lent, en utilisant l'approche de la perturbation singulière. Ils considèrent que les faibles interactions perturbent le système. Ainsi, Les auteurs dans (Altman et al., 2004a), étudient la perturbation singulière et régulière des chaînes de Markov sur un espace d'états dénombrable. Ils considèrent que les probabilités de transition entre les classes ergodiques perturbent le système (transitions rares) pour appliquer la perturbation singulière. Ils trouvent une représentation de la distribution stationnaire sous forme d'une série de Taylor en fonction du paramètre de la perturbation. Puis, ils étendent leurs résultats vers le processus quasi-naissance et mort dans les files d'attente.

Notre approche est basée sur la modélisation du système par une chaîne de Markov à deux dimensions, dont le premier correspond au nombre d'appels temps réel et le deuxième au nombre d'appels *best effort*. Afin d'obtenir la distribution stationnaire d'états d'équilibre approximativement, nous utilisons la perturbation singulière avec un processus quasi de naissance et de mort QBD (Latouche et Ramaswami, 1999). Nous considérons deux échelles de temps : une rapide et l'autre lente. Le premier représentant le nombre d'appels RT change rapidement et le deuxième représentant le changement d'appels BE lentement. La perturbation singulière nous a permis de représenter les probabilités d'états d'équilibre comme une série de Taylor. Les termes de cette série donnent une approximation de la distribution stationnaire que nous avons trouvé

par l'analyse spectrale dans le premier chapitre. Le premier terme de la série donne déjà une très bonne approximation quand le système identifie bien les deux échelles de temps citées auparavant.

3.2 Formulation du problème

3.2.1 Hypothèses

Les métriques de performances (comme le débit, le temps de séjour et la probabilité de blocage) sont très difficiles à calculer explicitement. Pour cela, dans le chapitre 2, nous avons utilisé l'analyse spectrale pour obtenir la forme explicite de la distribution stationnaire. Certes, la solution exacte trouvée dans le chapitre précédent, nous a permis d'évaluer les performances du système en fonction de certains paramètres du système. Mais il reste difficile à calculer la distribution stationnaire et surtout quand le nombre d'appels RT augmente, car les performances trouvées par l'analyse spectrale dépendent directement de ceux en RT. Donc, quand le nombre d'appels RT augmente, le système va prendre un temps important pour le calcul des performances et s'il y a des erreurs dû aux interférences des évanouissements, d'effet de masque, etc., le système doit refaire tout le calcul mathématique. Ce qui va rendre le système plus lourd au niveau temps de calcul.

Dans ce chapitre, nous introduisons une approche simple et rapide permettant d'évaluer en temps réel les performances du système. L'approche est basée sur l'hypothèse suivante : Dans le système WCDMA supportant les services RT et BE, nous supposons que selon les statistiques du canal, les appels BE restent longtemps durant leurs services (comme le cas du service Chat) et ceux en RT prennent généralement moins de temps de service (comme le cas du service voix). Donc, nous construisons deux groupes d'appels dans le système, un représente les transitions rapides causées par les appels RT et l'autre représente les faibles transitions causées par les appels BE. Ensuite, nous formulons mathématiquement une approximation de ces transitions. Ces approximations touchent la matrice des taux de transitions (\mathbf{Q}) de la chaîne de Markov décrivant le système, le taux d'arrivée d'appels BE (λ_{BE}) et le taux de service μ_{BE} . Nous définissons ces approximations comme suit

$$\mathbf{Q}(\epsilon) = \mathbf{Q}_0 + \epsilon\mathbf{Q}_1, \quad (3.1)$$

$$\lambda_{BE} = \epsilon\bar{\lambda}_{BE}, \quad (3.2)$$

$$\mu_{BE} = \epsilon\bar{\mu}_{BE}, \quad (3.3)$$

où ϵ est un paramètre strictement positif très petit représentant les faibles transitions, $\bar{\lambda}_{BE}$ est un paramètre strictement positif lié aux taux des arrivées, $\bar{\mu}_{BE}$ est un paramètre strictement positif dépendant des taux de services, $\mathbf{Q}(\epsilon)$ est un générateur correspondant à la chaîne de Markov perturbée; \mathbf{Q}_0 est un générateur non perturbé qui correspond à des interactions fortes et $\epsilon\mathbf{Q}_1$ est le terme perturbé correspond aux interactions

faibles. Nous avons les structures matricielles suivantes

$$\mathbf{Q}_0 = \begin{bmatrix} \mathbf{Q}_{RT} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{Q}_{RT} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{RT} & \dots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots \end{bmatrix},$$

$$\mathbf{Q}_1 = \begin{bmatrix} -\mathbf{A}_0 & \mathbf{A}_0 & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{A}_2 & -\mathbf{A}_2 - \mathbf{A}_0 & \mathbf{A}_0 & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{A}_2 & -\mathbf{A}_2 - \mathbf{A}_0 & \mathbf{A}_0 & \dots \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \ddots \end{bmatrix}.$$

où \mathbf{Q}_0 a une infinité de blocs fini \mathbf{Q}_{RT} et \mathbf{A}_0 et \mathbf{A}_2 sont des matrices carrées d'ordre $M_{RT} + 1$, tel que

$$\begin{aligned} \mathbf{A}_0 &= \text{diag}(\bar{\lambda}_{BE}), \\ \mathbf{A}_2 &= \text{diag}(\nu(i); \quad i = 0, 1, 2, \dots, M_{RT}). \end{aligned}$$

Ainsi, la matrice \mathbf{Q}_{RT} est carrée d'ordre $M_{RT} + 1$ et représente les arrivées et les départs d'appels RT

$$\begin{aligned} \mathbf{Q}_{RT}[i, i-1] &= i\mu_{RT}, \\ \mathbf{Q}_{RT}[i, i] &= -(\lambda_{RT} + i\mu_{RT}), \quad i = 0, \dots, M_{RT} - 1, \\ \mathbf{Q}_{RT}[M_{RT}, M_{RT}] &= -M_{RT}\mu_{RT}, \\ \mathbf{Q}_{RT}[i, i+1] &= \lambda_{RT}. \end{aligned}$$

Soit $\pi^{(\epsilon)}(i, j) \geq 0$ ($i = 0, \dots, M_{RT}$ et $j \geq 0$) la distribution de probabilité des états du nombre d'appels RT et BE à l'équilibre. Puisque $\mathbf{Q}(\epsilon)$ est irréductible ($\epsilon > 0$), alors cette probabilité invariante de mesure existe, unique et satisfait le système suivant :

$$\pi^{(\epsilon)}\mathbf{Q}(\epsilon) = \mathbf{0}, \quad (3.4)$$

$$\pi^{(\epsilon)}\mathbf{1} = 1, \quad (3.5)$$

où $\mathbf{1}$ est un vecteur unité de même dimension que le vecteur $\pi^{(\epsilon)}$. La chaîne de Markov a un espace d'états

$$E = \{(i, j), 0 \leq i \leq M_{RT}, j \geq 0\},$$

Qu'on peut partitionner en sous ensembles $l(j)$ tels que :

$$F = \{l(0), l(1), l(2), \dots\},$$

où $l(j) = \{(0, j), \dots, (M_{RT}, j)\}$ pour $j \geq 0$.

La matrice \mathbf{Q}_0 contient plusieurs classes ergodiques $l(j)$, chacune associée à un générateur \mathbf{Q}_{RT} qui est un bloc de la matrice diagonale \mathbf{Q}_0 .

Remarque 1 Nous signalons que, notre analyse est autant appliquée, dans le cas où les appels RT se produisent plus fréquemment que dans le cas de ceux en BE. Nous identifions dans ce cas, deux échelles d'interactions, une représente l'interaction rapide d'appels RT à intra-groupes et l'autre représente l'interaction moins fréquente inter-groupes d'appels BE. D'où l'approximation de ces interactions formulée en fonction des taux de transitions

$$\mathbf{Q}(\epsilon) = \frac{1}{\epsilon} \mathbf{Q}_0 + \mathbf{Q}_1, \quad (3.6)$$

$$\lambda_{RT} = \frac{1}{\epsilon} \bar{\lambda}_{RT}, \quad (3.7)$$

$$\mu_{RT} = \frac{1}{\epsilon} \bar{\mu}_{RT}, \quad (3.8)$$

où $\frac{1}{\epsilon}$ exprime la forte interaction d'appels RT qui perturbent le système, $\bar{\lambda}_{RT}$ est un paramètre strictement positif lié aux taux des arrivées, $\bar{\mu}_{RT}$ est un paramètre strictement positif dépendant des taux de services.

L'objectif de la section ci-après, est de trouver la représentation analytique de la distribution stationnaire $\pi(\epsilon)$, de la chaîne de Markov perturbée $\mathbf{Q}(\epsilon)$, en fonction du paramètre de la perturbation ϵ .

3.3 Série de Taylor

3.3.1 Définitions et Hypothèses

Premièrement, nous avons besoin d'introduire quelques définitions de base. Soit un vecteur fixe dénombrable δ de la forme $\delta = (\delta_0, \delta_1, \dots)$ tel que $\delta_i \geq 1$. Le vecteur δ et l'opérateur associé ont la même dimension. Notons $I_{M_{RT}} = \{0, 1, \dots, M_{RT}\}$ et $I_\infty = \{0, 1, \dots\}$, deux sous ensembles, alors la norme δ -norme est définie selon l'opérateur associé comme suit :

1. La norme δ -norme du vecteur v est définie comme suit :

$$\|v\|_\delta = \max_{i \in I} \frac{|v_i|}{\delta_i}, \text{ où } v_i = (v_0, v_1, \dots).$$

2. La norme δ -norme de l'opérateur fini $B = (b_{ij})$ est donnée par

$$\|B\|_\delta = \max_{i \in I_{M_{RT}}} \frac{\sum_{j=0}^{M_{RT}} |b_{ij}| \delta_j}{\delta_i}.$$

3. L'opérateur $D = (d_{jk})$ d'infinité de blocs (d_{jk}) est donnée par la formule suivante :

$$\|D\|_\delta = \max_{j \in I_\infty} \frac{\sum_{k \in I_\infty} \|d_{jk}\|_\delta \delta_k}{\delta_j}.$$

La chaîne de Markov doit vérifier les hypothèses du lemme suivant :

- Lemme 6**
1. La chaîne de Markov imperturbée a plusieurs classes ergodiques. Ces classes sont $l(j)$ pour tout $j \in I_\infty$.
 2. Chaque sous chaîne de Markov qui correspond à une classe ergodique de \mathbf{Q}_0 est uniformément stable au sens de Lyapunov.
 3. La chaîne de Markov agrégée est irréductible et stable au sens de Lyapunov.
 4. Le générateur \mathbf{Q}_1 est δ -borné ($\|\mathbf{Q}_1\|_\delta < \infty$).

Démonstration du Lemme 6 1. A partir des analyses précédentes, la chaîne de Markov non perturbée est constituée de plusieurs classes $l(j)$ pour $j \in I_\infty$ et il n'y a pas d'états de transit. Or, puisque les valeurs λ_{RT} et μ_{RT} sont positives, alors la classe $l(j)$ est une classe finie ergodique.

2. L'opérateur de transition de la chaîne de Markov imperturbée est $\mathbf{P}^{(0)} = r_0^{-1}\mathbf{Q}_0 + \mathbf{I}$ dont la forme est

$$\mathbf{P}^{(0)} = \mathbf{I} + \begin{pmatrix} r_0^{-1}\mathbf{Q}_{RT} + \mathbf{I} & \mathbf{0} & \dots \\ \mathbf{0} & r_0^{-1}\mathbf{Q}_{RT} + \mathbf{I} & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix},$$

où \mathbf{I} est une matrice unité d'ordre $M_{RT} + 1$, \mathbf{I} matrice unité par bloc et $r_0^{-1} = \frac{1}{\lambda_{RT} + M_{RT}\mu_{RT}}$.

Posons $\mathbf{P}_{RT} = \mathbf{I} + r_0^{-1}\mathbf{Q}_{RT}$ matrice de probabilité de transition. Puisque nous avons $P_{RT}(0,0) = 1 - \frac{\lambda_{RT}}{r_0} > 0$, alors la chaîne correspond aux classes ergodiques de la chaîne de Markov non perturbée, est fortement apériodique, ce qui implique qu'il existe des constantes positives c et β telles que $c > 0$, $0 < \beta < 1$ (voir (Altman et al., 2004a) pour plus de détails).

$$\|\mathbf{P}_{RT}^n - \mathbf{1}\tilde{q}_{l(i)}\|_\delta \leq c\beta^n, \forall n \geq 0 \text{ et } \beta < 1.$$

3. Soit $\tilde{\mathbf{Q}}_1$ un générateur de la chaîne de Markov agrégée sur l'espace F où chaque groupe $l(j)$ est remplacé par un état agrégé j (Par simplicité, nous remplaçons $l(j)$ par j).

$$(\tilde{\mathbf{Q}}_1)_{i,j} = \tilde{q}_i \mathbf{Q}_1[i,j] \underline{\xi}, \quad \forall (i,j) \in I_\infty \times I_\infty,$$

où $\underline{\xi} = (1, \dots, 1)^t \in \mathfrak{R}^{(M_{RT}+1) \times 1}$. Cette matrice agrégée représente les transions entre les groupes. Chaque colonne dans $\tilde{\mathbf{Q}}_1$ correspond à un groupe dans F . La matrice $\tilde{\mathbf{Q}}_1$ est tridiagonale :

$$\begin{aligned} \tilde{\mathbf{Q}}_1[j, j+1] &= \tilde{q}_j \mathbf{A}_0 \underline{\xi} \\ \tilde{\mathbf{Q}}_1[j, j-1] &= \tilde{q}_j \mathbf{A}_2 \underline{\xi} \\ \tilde{\mathbf{Q}}_1[j, j] &= -\tilde{q}_j (\mathbf{A}_0 + \mathbf{A}_2) \underline{\xi}, (\forall j > 0) \\ \tilde{\mathbf{Q}}_1[0, 0] &= -\tilde{q}_0 (\mathbf{A}_0) \underline{\xi} \end{aligned}$$

où $\tilde{q}_j \mathbf{A}_n \underline{\xi} = \sum_{i=0}^{M_{RT}} \tilde{q}_j(i) (\mathbf{A}_n)_{(i,i)}$ pour $n = 0, 2$. La distribution invariante \bar{v} correspondant à ce générateur, est obtenue en résolvant le système d'équations suivant

$$\bar{v}\tilde{\mathbf{Q}}_1 = \mathbf{0}, \sum_{i=0}^{\infty} \bar{v}_i = 1 \quad (3.9)$$

où $\bar{v} = [\bar{v}_0, \bar{v}_1, \dots]$ est le vecteur de probabilité $\tilde{q}_j = [\tilde{q}_j(0), \dots, \tilde{q}_j(M_{RT})]$ est une solution unique de la forme

$$\tilde{q}_j \mathbf{Q}_{RT} = 0, \tilde{q}_j \underline{\xi} = 1, \quad (3.10)$$

Toute fois, le vecteur \tilde{q}_j est donné par

$$\tilde{q}_j(n) = \frac{(\rho_{RT})^n}{n!} \left(\sum_{i=0}^{M_{RT}} \frac{(\rho_{RT})^i}{i!} \right)^{-1}, 0 \leq n \leq M_{RT} \quad (3.11)$$

Le générateur $\tilde{\mathbf{Q}}_1$ devient

$$\tilde{\mathbf{Q}}_1 = \begin{bmatrix} -\bar{\lambda}_{BE} & \bar{\lambda}_{BE} & 0 & \dots \\ c & -\bar{\lambda}_{BE} - c & \bar{\lambda}_{BE} & \ddots \\ 0 & c & -\bar{\lambda}_{BE} - c & \bar{\lambda}_{BE} \\ 0 & \ddots & \ddots & \ddots \end{bmatrix},$$

où

$$c = \sum_{i=0}^{N_{RT}} v(i)\tilde{q}(i) + \sum_{i=N_{RT}+1}^{M_{RT}} v(i)\tilde{q}(i)$$

Posons

$$\alpha(M_{RT}) = \frac{\bar{\lambda}_{BE}}{\sum_{i=0}^{N_{RT}} v(i)\tilde{q}(i) + \sum_{i=N_{RT}+1}^{M_{RT}} v(i)\tilde{q}(i)}.$$

Le générateur de la chaîne de Markov agrégée correspond au générateur d'un processus QBD homogène du modèle M/M/1 dont le taux d'arrivée est $\bar{\lambda}_{BE}$ et le taux de départ est c . Nous pouvons vérifier que la chaîne de Markov agrégée est irréductible ergodique et fortement apériodique si $\alpha(M_{RT}) < 1$.

Soient $\tilde{\mathbf{P}} = r_{BE}^{-1}\tilde{\mathbf{Q}}_1 + \mathbf{I}$, la matrice des probabilités de transition associée à la chaîne de Markov agrégée et $r_{BE}^{-1} = \frac{1}{\bar{\lambda}_{BE} + 2v(0)}$. Nous remarquons que l'état $(0,0)$ est fortement apériodique, car la probabilité suivante à cet état est strictement positive $\tilde{\mathbf{P}}(0,0) = 1 - \frac{\bar{\lambda}_{BE}}{r_0} > \frac{v(0)}{r_0} > 0$, alors la fonction de Lyapunov δ est vérifiée $\tilde{\mathbf{P}}\delta \leq \gamma\delta + b\mathbf{1}_{(0,0)}$, où $b = \sqrt{c}(1 - \sqrt{\alpha(M_{RT}))}\delta_0, \delta_{l(i)} = (\sqrt{\alpha(M_{RT}))}^i$ et $\gamma = 1 - (\bar{\lambda}_{BE} - c)^2$ (pour plus de détail voir (Altman et al., 2004a; Meyn et Tweedie, 1994)).

Nous rappelons que les états de la chaîne agrégée sont les classes ergodiques de la chaîne originale, en présence de la perturbation, elle représente exactement les faibles transitions entre les groupes.

4. La norme de l'opérateur \mathbf{Q}_1 est majorée par une constante positive, car nous avons

$$\|\mathbf{Q}_1\|_\delta \triangleq \max_{j \in I_\infty} \frac{\sum_{k \in I_\infty} \|\mathbf{Q}_1(i, k)\| \delta_k}{\delta_j} = \max_{i \in I_\infty} \left\{ \frac{\|\mathbf{A}_2(i, \cdot)\|_\delta (\delta_0 + \delta_1)}{\delta_0}, \right. \\ \left. \frac{\|\mathbf{A}_0\|_\delta \delta_{i-1} + \|(\mathbf{A}_2 + \mathbf{A}_0)\|_\delta \delta_i + \|\mathbf{A}_0\|_\delta \delta_{i+1}}{\delta_{i-1}} \right\},$$

où

$$\|\mathbf{A}_0\|_\delta = \bar{\lambda}_{BE} \|I\|_\delta = \bar{\lambda}_{BE}, \text{ et } \|\mathbf{A}_2\|_\delta = v(0).$$

Puisque nous avons

$$\max(v(i)) = v(0), \text{ et } \|(\mathbf{A}_2 + \mathbf{A}_0)\|_\delta = \bar{\lambda}_{BE} + v(0)$$

Alors nous obtenons

$$\|\mathbf{Q}_1\|_\delta = \max_{i \in I_{MRT}} \left\{ \bar{\lambda}_{BE} \left(1 + \frac{\delta_1}{\delta_0}\right), \bar{\lambda}_{BE} \left(1 + \frac{\delta_i}{\delta_{i-1}}\right) + v(0) \left(1 + \frac{\delta_{i+1}}{\delta_i}\right) \right\}.$$

Par conséquent, le générateur \mathbf{Q}_1 est δ -borné.

3.3.2 Représentation analytique de la distribution stationnaire

Nous présentons maintenant la distribution stationnaire des états d'équilibre de la chaîne de Markov perturbée, par une fonction analytique de paramètre ϵ . A partir du lemme 6, les hypothèses du Théorème 4.1 dans (Altman et al., 2004a) sont satisfaites, nous obtenons les résultats suivants

Théorème 4 La forme analytique de la probabilité de mesure invariante $\pi^{(\epsilon)}$ en fonction du paramètre ϵ est

$$\pi^{(\epsilon)} = \pi^{(0)} + \pi^{(1)}\epsilon + \pi^{(2)}\epsilon^2 + \pi^{(3)}\epsilon^3 + \dots \quad (3.12)$$

avec $0 < \epsilon \leq \min(\epsilon_{max}, \frac{1}{\|\mathbf{U}\|_\delta})$ et la norme $\|\mathbf{U}\|_\delta$ est donnée par la proposition 3.3.3.

Les coefficients probabilistiques de $\pi^{(\epsilon)}$ peuvent être calculés d'une manière recursive par les formules suivantes

$$\pi^{(0)} \mathbf{Q}_0 = \mathbf{0}, \quad (3.13)$$

$$\sum_{(i=0, j=0)}^{(M_{RT}, \infty)} \pi^{(0)}(i, j) = 1. \quad (3.14)$$

$$\pi^{(n)} \mathbf{Q}_0 + \pi^{(n-1)} \mathbf{Q}_1 = \mathbf{0}, \quad (3.15)$$

$$\sum_{(i=0, j=0)}^{(M_{RT}, \infty)} \pi^{(n)}(i, j) = \mathbf{0}, \quad \forall n \geq 1. \quad (3.16)$$

Tous les états de la chaîne de Markov non perturbée ne communiquent pas entre eux. Pour tout $\epsilon > 0$ suffisamment petit, les états de la chaîne de Markov perturbée communiquent entre eux, c'est à dire qu'il y a une discontinuité de $\pi(\epsilon)$ à $\epsilon = 0$. Donc la distribution stationnaire de la chaîne non perturbée n'est pas unique, en particulier, elle s'écrit en fonction des solutions stationnaires de toutes les classes ergodiques. Par exemple nous avons :

$$\pi_j^{(0)} = \bar{v}_j \tilde{q}_j,$$

où \bar{v}_j est la solution de l'équation (3.9) et sa forme est :

$$\bar{v}_i = \alpha(M_{RT})^i \bar{v}_0, \text{ avec } \bar{v}_0 = 1 - \alpha(M_{RT}), \forall i \in I_\infty \quad (3.17)$$

Pour trouver la forme explicite de la distribution stationnaire du système, nous introduisons quelques variables. Soit $\xi_{l(i)}$ le vecteur propre colonne, qui correspond à la valeur propre 1 de la matrice de transition d'une chaîne de Markov imperturbée, qui fournit la probabilité que le processus soit absorbé par la classe $l(i)$, sachant qu'il commence à partir de l'état i ($\mathbf{Q}_0 \xi_{l(i)} = 0$). Dans notre modèle, l'expression de ce vecteur est donnée pour $\forall j \in I_{M_{RT}}$ et $\forall k \in I_\infty$ par :

$$\xi_{l(i)}(k, j) = \begin{cases} 1 & \text{si } k = i; \\ 0 & \text{sinon.} \end{cases}$$

Soient \mathbf{N} la matrice des probabilités d'absorption par des classes ergodiques, c'est à dire nous avons $\mathbf{N} = [\xi_{l(0)} \quad \xi_{l(1)} \quad \dots]$ et \mathbf{M} matrice des probabilités stationnaires de la chaîne de Markov \mathbf{Q}_0 dont la forme est $\mathbf{M} = \text{diag}(\tilde{q}_{l(0)}, \tilde{q}_{l(1)}, \dots)$.

Proposition 3.3.3 *La solution d'équations (3.13)-(3.15) en utilisant les conditions de normalisation est donnée de manière récursive pour :*

1. Le premier coefficient, par

$$\pi_{(i,j)}^{(0)} = (1 - \alpha(M_{RT})) \alpha(M_{RT})^j \frac{(\rho_{RT})^i}{i!} \frac{1}{\sum_{k=0}^{M_{RT}} \frac{(\rho_{RT})^k}{k!}}. \quad (3.18)$$

2. Le coefficient d'ordre n ($n > 0$), par

$$\pi^{(n)} = \pi^{(0)} \mathbf{U}^n, \mathbf{U} = \mathbf{Q}_1 \mathbf{H} [\mathbf{I} + \mathbf{Q}_1 \Lambda \tilde{\mathbf{H}} \mathbf{M}], \quad (3.19)$$

avec

$$\|\mathbf{U}\|_\delta = g_1 \|\mathbf{H}\|_\delta (1 + g_1 \|\Lambda \tilde{\mathbf{H}} \mathbf{M}\|_\delta).$$

Les matrices $\mathbf{H}^0, \tilde{\mathbf{H}}$ sont des matrices de déviation de la chaîne imperturbée et de la chaîne agrégée, respectivement. Aussi nous avons

$$\Lambda = \begin{pmatrix} \underline{\xi} & \underline{0} & \underline{0} & \underline{0} & \dots \\ \underline{0} & \underline{\xi} & \underline{0} & \underline{0} & \dots \\ \underline{0} & \underline{0} & \underline{\xi} & \underline{0} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \text{ avec } \underline{\xi} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \text{ et } \underline{0} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}.$$

Ainsi la matrice \mathbf{H} est donnée par

$$\mathbf{H} = [-\mathbf{Q}_0 + \mathbf{NM}]^{-1} - \mathbf{NM} = \text{diag}(\mathbf{H}_0, \dots)$$

où

$$\mathbf{H}_0 = [q^* - \mathbf{Q}_{RT}]^{-1} - q^*, \quad q^* = \underline{\zeta} \tilde{q},$$

Car nous avons $\tilde{q} = \tilde{q}_j$, d'où le δ -norme de la matrice de déviation \mathbf{H}^0 est donnée par :

$$\|\mathbf{H}\|_\delta = \|H_0\|_\delta, \quad \tilde{\mathbf{H}} = [-\tilde{\mathbf{Q}}_1 + \tilde{\mathbf{Q}}_1^*]^{-1} - \tilde{\mathbf{Q}}_1^*, \quad \tilde{\mathbf{Q}}_1^* = \zeta_{\mathbf{B}} \tilde{q}, \quad \zeta_{\mathbf{B}} = (1, \dots, 1) \in \mathfrak{R}^{\infty \times 1}$$

Les blocs matricielles $U_{i,j}$ de la matrice $\mathbf{U} = \{U_{i,j}\}_{i,j \in F}$ sont obtenus par la relation :

$$U_{i,j} = \mathbf{Q}_{i,j}^1 \mathbf{H}_j + \sum_{l \in F} \mathbf{Q}_{i,l}^1 \mathbf{H}_l \sum_{k \in B} \mathbf{Q}_{l,k}^1 \tilde{\mathbf{H}}_{k,j} \mathbf{1} q_j.$$

Nous constatons ici que notre approche de la perturbation singulière est plus simple que l'approche d'analyse spectrale que nous avons utilisé dans (Ibrahimi et al., 2007). Nous avons juste besoin de calculer la distribution stationnaire pour chaque classe récurrente de la chaîne de Markov non perturbée \mathbf{Q}_0 dans (3.10). Aussi on calcule la distribution invariante de la chaîne agrégée $\tilde{\mathbf{Q}}_1$ par (3.17).

Lemme 7 *La distribution stationnaire du nombre d'appels en BE selon l'ordre des coefficients de la série de Taylor est donnée en distinguant les deux cas suivants :*

1. Pour le premier coefficient $\pi^{(0)}$, nous trouvons

$$P_\epsilon^{(0)}(Y = j) = \pi_j^{(0)} \mathbf{1} = (1 - \alpha(M_{RT})) \alpha(M_{RT})^j.$$

2. Pour les autres coefficients $\pi^{(n)}$ ($n \geq 1$), nous trouvons

$$P_\epsilon^{(n)}(Y = j) = \pi_j^{(0)} \mathbf{1} + \epsilon \pi_j^{(1)} \mathbf{1} + \dots + \epsilon^n \pi_j^{(n)} \mathbf{1},$$

$$\text{où } \pi_j^{(n)} = \sum_{i \in B} \pi_i^{(0)} \mathbf{u}_{i,j}^{(n)} \text{ et } \mathbf{u}^n = \{\mathbf{u}_{i,j}^{(n)}\}.$$

Après avoir trouvé les formes explicites des probabilités marginales, nous déduisons certains métriques de performance par la suite.

3.3.4 Temps de séjour d'appels *best effort*

Maintenant, nous pouvons déterminer les métriques de performance du système de manière approximative. En effet, le nombre moyen marginal $E_\epsilon(Y)$ d'appels BE dans le système est donné par

$$E_\epsilon(Y) = \sum_{j=0}^{\infty} j P_\epsilon^{(n)}(Y = j), \quad n \geq 0. \quad (3.20)$$

Le temps de séjour moyen d'appels BE dans le système est donné par la formule de Little

$$T_\epsilon^{BE} = \frac{1}{\lambda_{BE}} \sum_{j=0}^{\infty} j P_\epsilon^{(n)}(Y = j), \quad n \geq 0. \quad (3.21)$$

En particulier, pour le premier coefficient de la série de Taylor, c-à-d pour $n = 0$, nous trouvons les performances suivantes

$$\begin{aligned} E_\epsilon(Y) &= \frac{\alpha(M_{RT})}{1 - \alpha(M_{RT})}, \\ T_\epsilon^{BE} &= \frac{1}{\lambda_{BE}} \frac{\alpha(M_{RT})}{1 - \alpha(M_{RT})}. \end{aligned}$$

Nous écrivons le temps de séjour moyen d'appels BE en fonction du seuil L_{BE} , sous la forme finale

$$T^{BE}(L_{BE}) = \frac{\frac{1}{\epsilon}}{\gamma[\sum_{i=0}^{N_{RT}} \tilde{q}(i)\beta_{\epsilon'} + \frac{L_{BE}}{1-L_{BE}} \sum_{i=N_{RT}+1}^{M_{RT}} \tilde{q}(i)] - \bar{\lambda}_{BE}}, \quad (3.22)$$

où $\gamma = \frac{\bar{\mu}_{BE}W}{\Gamma E_{BE}/N_0}$ et $\beta_{\epsilon'} = \frac{\Theta - i\Delta_{RT}^r}{1 - \Theta + i\Delta_{RT}^r}$. L'équation ci-dessus montre l'influence du seuil L_{BE} sur le temps de séjour d'appels BE. Nous rappelons que les nombres d'appels M_{RT} et N_{RT} dépendent aussi du seuil L_{BE} .

Remarque 2 Nous signalons que le choix de l'ordre de l'approximation dans la série de Taylor dépend du degré de précision entre les deux approches la perturbation singulière et l'analyse spectrale. Donc, si le premier coefficient ne donne pas une bonne approximation de la distribution stationnaire, sachant que l'hypothèse considérée, est toujours valide, nous passons au deuxième coefficient et ainsi de suite jusqu'à ce qu'on trouve la bonne approximation.

3.4 Résultats numériques

Nous considérons les paramètres suivants. Le débit minimum d'appels RT est 4.75 *kbps* et le débit maximum est 12.2 *kbps*. L'énergie sur bruit pour transmettre un bit de type RT est $E_{RT}/N_0 = 4.1$ dB, $E_{BE}/N_0 = 4.1$ dB (Holma et Toskala, 2001), $\bar{p}_{BE} = 0.55$, $\bar{p}_{RT} = 0.5$, $R_T = 38$ *kbps* et $\Theta = 1 - 10^{-5}$. Le taux d'arrivée d'appels BE est $\lambda_{BE} = 0.209\epsilon$, leur taux de départ est $\mu_{BE} = 10^{-5}\epsilon$ et le paramètre de la perturbation est $\epsilon = 10^{-3}$. La figure 3.1, présente une comparaison de la probabilité marginale d'appels BE, obtenue par les deux méthodes : l'analyse spectrale (voir la première partie du chapitre 2) et la perturbation singulière. Nous avons considéré dans l'approche de perturbation le premier terme de la série de Taylor comme approximation de la distribution stationnaire. Nous remarquons que ce premier coefficient de Taylor donne une très bonne approximation quand le paramètre de la perturbation est suffisamment petit, par exemple $\epsilon = 10^{-3}$.

Les figures 3.3, 3.4 et 3.2, exposent le temps de séjour d'appels BE en fonction du seuil L_{BE} . Nous rappelons que l'hypothèse de l'applicabilité de la méthode perturbation singulière est que les appels BE se produisent moins fréquemment que les appels RT. Nous observons que le premier coefficient d'approximation dans la série de Taylor donne une parfaite approximation dans la figure 3.2. Nous constatons que l'erreur entre la Méthode exacte (analyse spectrale) et la méthode approximative (perturbation singulière)

devient significative quand le temps de service d'appels RT décroît (voir les figures 3.3 et 3.4). Ce Phénomène est due au fait que lorsque l'on utilise le premier terme de la série de Taylor, on obtient le comportement limite si ϵ tend vers zéro, c'est à dire, que les transitions du nombre d'appels RT se produisent beaucoup plus rapidement que les transitions du nombre d'appels BE. Toutefois, lorsque le temps de service d'appels RT diminue, les appels RT passent plus de temps dans le système, ce qui veut dire que les transitions du nombre d'appels RT devient de moins en moins lent et nous avons donc les mêmes configurations de transition pour les deux appels RT et BE. Or, ce n'est pas le cas dans notre hypothèse où nous avons supposé deux échelles de transitions différentes, une rapide (nombre d'appels RT qui se produisent rapidement) et une lente (nombre d'appels BE qui se produisent lentement). Néanmoins, le fournisseur de service ou l'opérateur peut utiliser cette approche basée sur l'approximation pour évaluer en temps réel avec un calcul rapide des métriques de performances d'appels RT et BE. Ainsi, le fournisseur peut décider dans ce cas entre la meilleure configuration du choix de seuil L_{BE} et le contrôle d'admission qui améliore l'utilisation de la bande passante. En effet, les performances du système dépendent de plusieurs facteurs : les taux d'arrivée d'appels, la durée moyenne d'appel RT, les effets de masques, Ces facteurs changent dynamiquement et le système a besoin de réévaluer les performances à nouveau. Par conséquent, notre approche permet à l'opérateur d'avoir un meilleur CAC et une meilleure utilisation de la bande passante tout en satisfaisant les demandes de qualité de service d'appels RT et BE.

Enfin, l'analyse des résultats nous permet de constater que l'erreur entre la solution exacte (méthode d'analyse spectrale) et la solution approximative (méthode de la perturbation singulière) croît quand la durée moyenne de service ($1/\mu_{RT}$) d'un paquet de type RT augmente. Cet écart d'erreur est dû au fait que la durée de service ($1/\mu_{RT}$) d'un paquet RT augmente alors que les appels en RT restent longtemps dans le système ; ce qui n'est pas le cas dans notre modélisation où nous avons supposé qu'ils varient rapidement dans le système. Il est donc normal que l'approximation soit distincte de la solution exacte.

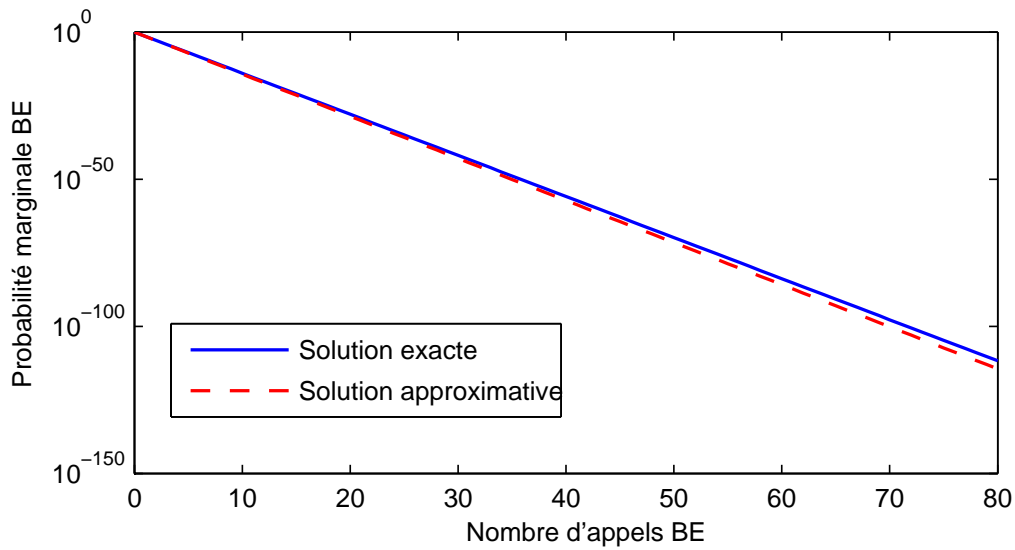


FIGURE 3.1 – Probabilité marginale en fonction du nombre d'appels BE, pour $\epsilon = 10^{-3}$, $\lambda_{BE} = 0.209\epsilon$, $\mu_{BE}^\epsilon = 10^{-5}\epsilon$, $\lambda_{RT} = 0.1557$, $L_{BE} = 0.752$ et $\mu_{RT} = 0.1$. Comparaison entre deux méthodes : Analyse Spectrale (solution exacte) et Perturbation Singulière (solution approximative).

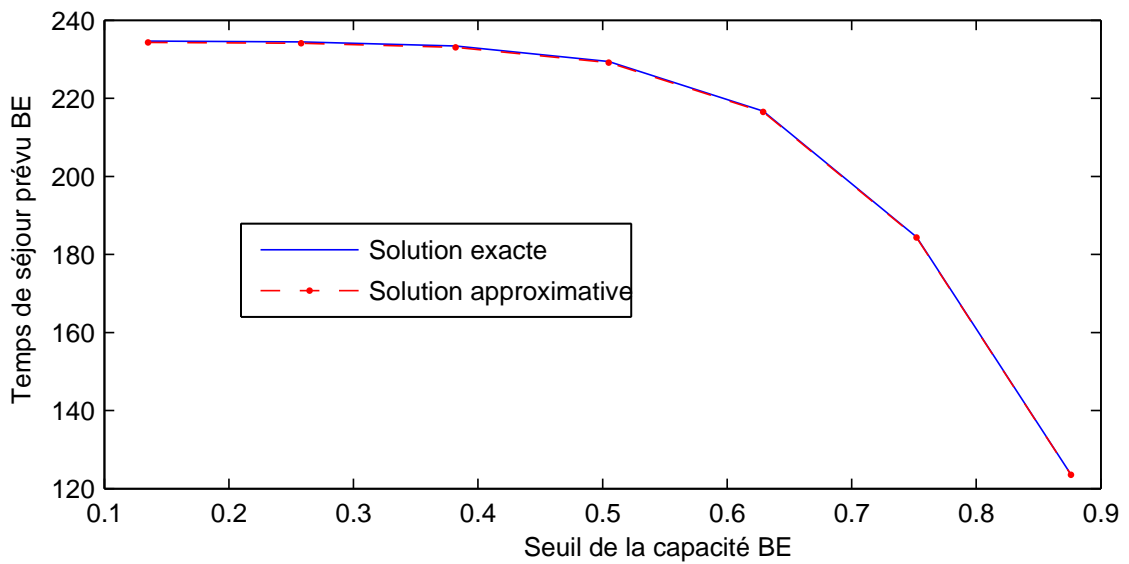


FIGURE 3.2 – Temps de séjour d'appels BE en fonction du seuil L_{BE} , pour $\epsilon = 10^{-3}$, $\lambda_{BE} = 0.209\epsilon$, $\mu_{BE}^\epsilon = 10^{-5}\epsilon$, $\lambda_{RT} = 0.1557$ et $\mu_{RT} = 100$. Comparaison entre deux méthodes : Analyse Spectrale (solution exacte) et Perturbation Singulière (solution approximative).

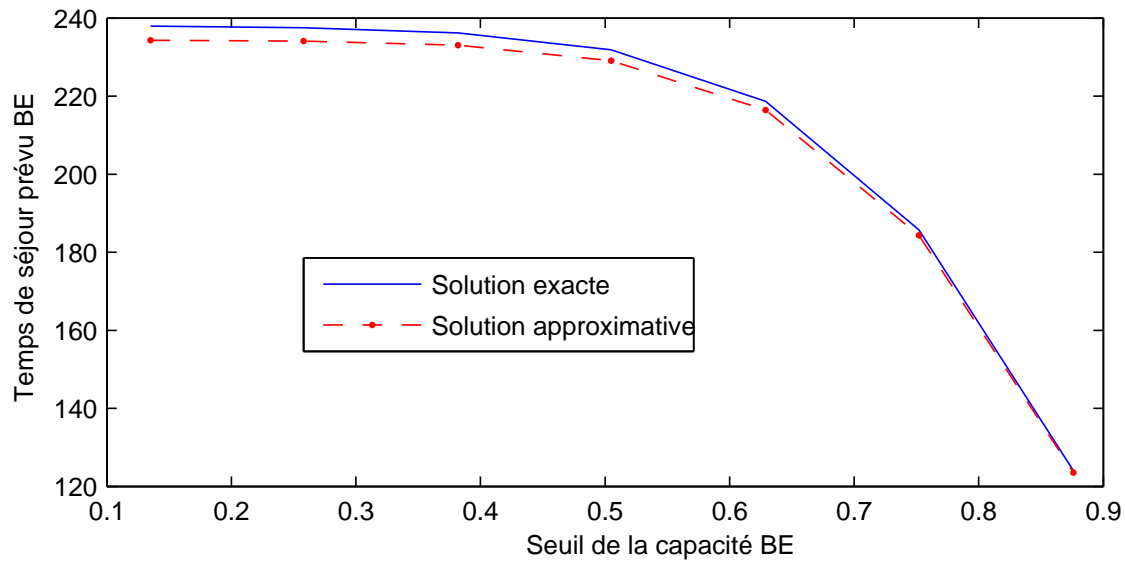


FIGURE 3.3 – Temps de séjour d’appels BE en fonction du seuil L_{BE} , pour $\epsilon = 10^{-3}$, $\lambda_{BE} = 0.209\epsilon$, $\mu_{BE}^\epsilon = 10^{-5}\epsilon$, $\lambda_{RT} = 1.5573$ et $\mu_{RT} = 1$. Comparaison entre deux méthodes : Analyse Spectrale (solution exacte) et Perturbation Singulière (solution approximative).

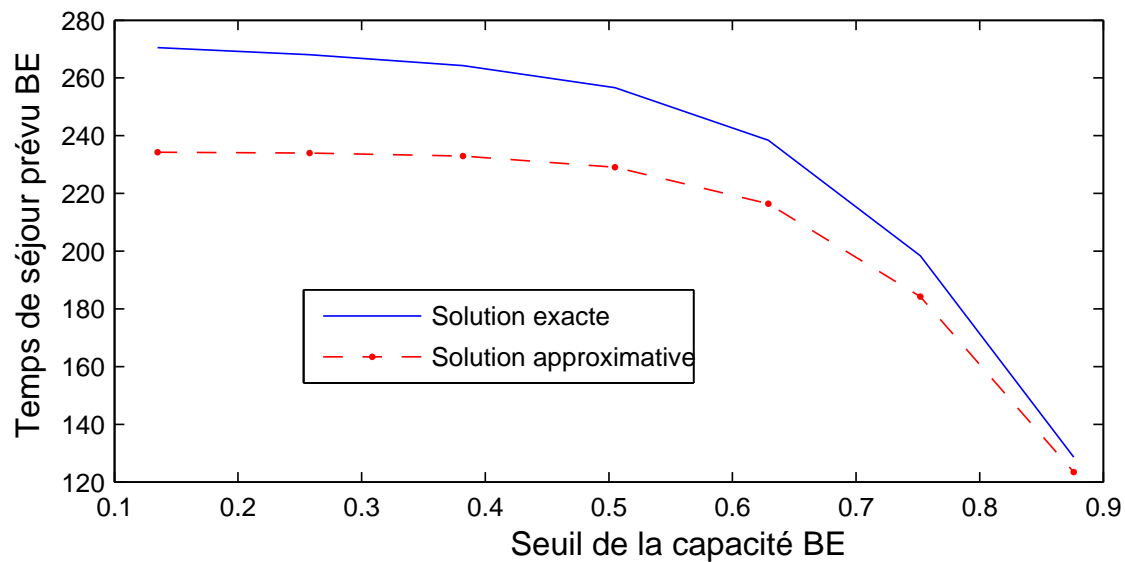


FIGURE 3.4 – Temps de séjour d’appels BE en fonction du seuil L_{BE} , pour $\epsilon = 10^{-3}$, $\lambda_{BE} = 0.209\epsilon$, $\mu_{BE}^\epsilon = 10^{-5}\epsilon$, $\lambda_{RT} = 0.1557$ et $\mu_{RT} = 0.1$. Comparaison entre deux méthodes : Analyse Spectrale (solution exacte) et Perturbation Singulière (solution approximative).

3.5 Conclusion

Dans le chapitre précédent, nous avons fourni une méthode exacte qui donne des formes explicites aux performances du système, en terme de probabilité d'existence et de blocage d'appels, et de temps de séjour. L'avantage de cette méthode est la réduction de la complexité du problème. Les performances du système ne dépendent que du nombre d'appels RT. Cette indépendance en nombre d'appels BE nous permet de ramener un problème avec un nombre arbitraire d'appels BE (voir infini) à un problème relativement simple. Alors quand le nombre d'appels RT augmente, la méthode prend un temps de calcul important et donne quelques erreurs, ce qui nous suggère à utiliser une autre approche complémentaire et rapide au niveau calcul. En effet, dans ce chapitre, nous avons étudié le système WCDMA avec deux types d'appels : temps réel (RT) qui ont des ressources dédiées et non temps réel ou *best effort* (BE) qui partagent la capacité restante du système. Nous avons considéré durant ce travail une capacité minimum réservée aux appels BE ainsi que toute portion de la capacité non utilisée par les appels RT. Notre approche d'analyse du système est basée sur la modélisation du système par une chaîne de Markov à deux dimensions, où la première correspond au nombre d'appels RT et la deuxième correspond au nombre d'appels BE dans le système. Néanmoins, nous avons utilisé une approximation basée sur l'approche de la perturbation singulière pour trouver une approximation simple de la distribution d'états d'équilibre de cette chaîne de Markov. Nous avons montré sous forme numérique que notre approche donne une bonne approximation de la distribution d'états d'équilibre avec des calculs plus simple que dans le cas de l'analyse spectrale. Toutefois, le système peut utiliser cette approche pour évaluer en temps réel les mesures de performances d'appels RT et BE.

Troisième partie

Gestion des ressources des réseaux mobiles HSDPA et IEEE802.16e

Chapitre 4

Gestion des ressources du réseau HSDPA en présence de la mobilité

Sommaire

4.1	Introduction	97
4.2	Formulation du problème	99
4.2.1	Modulation et codage adaptatif (AMC)	99
4.2.2	Bande passante par type d'appel	100
4.2.3	Modélisation	102
4.3	Mobilité intra cellule	102
4.3.1	Sans mobilité	102
4.3.2	Avec intra mobilité	103
4.3.3	Calcul des taux de migration	105
4.4	Mobilité inter cellules	106
4.5	Métriques des performances	107
4.6	Résultats numériques	109
4.7	Conclusion	117

4.1 Introduction

Dans la première partie de ce travail, nous avons analysé la capacité du lien montant du système WCDMA par deux approches : celle exacte (analyse spectrale), développée dans le chapitre 2 et celle approximative (perturbation singulière), développée dans le chapitre 3. Ces deux approches nous ont permis d'évaluer la capacité du système de manière exacte en fonction des paramètres du système (taux d'arrivée et de service d'appels ainsi que le nombre d'appels).

Or, pour mieux gérer la bande passante du système, il faut prendre en considération la variation de l'état du canal. Les utilisateurs qui sont proches de la station de base per-

çoivent généralement une bonne qualité du canal et donc une qualité de service satisfaisante. Par contre, ceux qui sont loin, leur qualité de service s'est vue dégradée. Puisque la technique WCDMA ne permet pas d'adapter la modulation selon la qualité du canal, nous nous intéressons dans ce chapitre à l'étude du problème de la variation de l'état du canal utilisant la technique HSDPA ; extension logicielle du système WCDMA. En effet, dans le réseau à haut débit HSDPA, le système adapte dynamiquement la qualité de service selon l'état du canal perçue par l'utilisateur. Cependant, quand la qualité du canal change de l'état bon vers l'état mauvais, l'utilisateur se trouve avec une qualité de service dégradée. Nous utilisons ainsi la technique de la modulation et du codage adaptatifs AMC (*Adaptive Modulation and Coding*). Ceci permet de maintenir un débit constant à l'utilisateur indépendamment de sa position et en tenant compte de sa mobilité.

Les auteurs dans (Li et Sampalli, 2007) développent un contrôle d'admission sur le lien descendant des réseaux à haut débit WiMAX et HSDPA en nous basant sur l'adaptation du lien radio de l'interface aire. Ils décomposent la cellule en nombre fini de régions prenant en compte un modèle simple de propagation opérant avec uniquement un *path loss* en fonction de la distance entre le mobile et la station de base. Ils considèrent dans leur étude une seule classe de service et modélisent le système par le réseau de file d'attente BCMP (Baskett et al., 1975; Chao et al., 2001) dont le nombre d'utilisateurs actifs est fixe. Les auteurs supposent que les taux d'arrivée d'appels dans chaque région dépendent uniquement du nombre moyen de mobiles qui se déplacent dans cette région, et que le temps moyen de service dépend seulement du temps que le mobile reste dans la région. Le travail de ces auteurs nous encourage à développer un nouveau mécanisme de contrôle d'admission qui prend en compte deux classes de service (temps réel et non temps réel), la mobilité intra et inter cellules. Aussi, considérons-nous le taux d'arrivée d'appels dans chaque région en fonction de la mobilité, du taux des nouvelles arrivées et de la durée moyenne de service. Le contrôle d'admission se base sur la subdivision de la cellule en un nombre fini de régions et chaque région possède une modulation différente. Il permet de garantir à chaque mobile son débit initial quelque soit sa position dans la cellule ou quand la qualité de son lien radio devient mauvaise. La décomposition de la cellule en régions a intéressé plusieurs chercheurs (Bonald et Proutière, 2003; Elayoubi et al., 2006). Pour évaluer les performances du système en termes de probabilité de blocage des nouveaux appels, de probabilité des pertes d'appels en cours et du débit moyen total, nous utilisons la théorie des files d'attentes de type BCMP (Baskett et al., 1975).

La différenciation de service entre les appels temps réel et non temps réel se fait en fonction du service demandé par l'utilisateur. La durée moyenne d'appels temps réel ne dépend pas des ressources consommées par le mobile et dépend uniquement de son comportement. Par contre, la durée moyenne de téléchargement de fichier pour les appels non temps réel dépend des ressources consommées par le mobile et de la taille moyenne de ces fichiers. La mobilité intra cellule est modélisée par le mouvement d'utilisateurs entre les régions (changement de la modulation au sein de la même station de base). La mobilité inter cellules est modélisée par le mouvement des utilisateurs entre les cellules (changement de la station de base). Nous avons constaté que la mobilité

inter cellules exige de notre mécanisme du CAC à prendre en compte les interférences inter cellules pour varier adaptativement la bande passante afin de maintenir un débit constant tout le temps. Donc le CAC assigne plus de bande passante pour chaque mobile et nous aurons moins d'appels dans le système, ce qui dégrade la capacité du système comme remarqué par (Viterbi et al., 1994). Ce mécanisme du CAC nous a permis de trouver un bon compromis entre la probabilité de perte d'un appel en cours et le blocage d'un nouveau appel, ainsi que le choix des ressources à réserver pour les appels en mobilité et une meilleure gestion des ressources libres.

4.2 Formulation du problème

4.2.1 Modulation et codage adaptatif (AMC)

Dans les réseaux mobiles, la qualité d'un signal reçu par un terminal dépend de plusieurs facteurs, comme la distance entre la station de base ou la perte de propagation due aux obstacles ou aux trajets multiples. Dans le but d'améliorer la performance d'un système, en termes de débit global et de fiabilité de la couverture radio, le signal transmis vers et par un utilisateur particulier est modifié de façon à prendre en compte les variations de la qualité du signal à travers un processus connu sous le nom d'adaptation au lien radio, ou de AMC (Modulation et Codage Adaptatif). L'AMC devient une approche standard dans les réseaux à haut débit comme HSDPA (Kwan et al., 2002; Nakamura et al., 2002). Comme indiqué dans (Qiu et al., 2002), l'idée derrière AMC est le changement dynamique de la modulation et du codage selon les conditions du canal radio. Si le rapport signal sur interférence plus le bruit (SNIR) est bon, le système peut affecter à l'utilisateur une forte modulation avec un taux de codage élevé (par exemple 64-QAM, 3/4 codage Turbo). Quand la qualité du canal se détériore, le système peut choisir une autre modulation plus faible avec un taux de codage faible sans perdre l'appel en cours, par exemple QPSK avec 1/2 codage Turbo. L'objectif du schéma de modulation et codage adaptatif (MCS) est de maintenir un taux d'erreur binaire (BER, *Bit Error Rate*) acceptable qui devrait être inférieur à un seuil donné BER_{th} pour tous les utilisateurs de la cellule. Soit $\{M_1, \dots, M_r\}$ un ensemble de r schémas de modulations et de codages adaptatifs. Selon la valeur du SINR, chaque utilisateur reçoit une modulation et un taux de codage tel que $BER \leq BER_{th}$. Les points d'intersection entre les valeurs moyennes du SINR et la modulation sont considérés comme des valeurs seuils, comme, $\{\gamma_r, \dots, \gamma_1, \gamma_0 = \infty\}$. Ces valeurs nous permettent de partitionner le SINR en r intervalles (régions) : $[\gamma_i, \gamma_{i-1})$, pour $i = 1, \dots, r$ (voir le tableau 4.1). Le k ème schéma MCS (M_k), est affecté à la région $[\gamma_k, \gamma_{k-1})$ si les conditions suivantes sont valides

$$BER(M_k, \gamma_k) \leq BER_{th} \text{ et } BER(M_k, \gamma_{k+1}) > BER_{th}.$$

Nous subdivisons la cellule en r régions concentriques de rayon R_i , $i = 1, \dots, r$, et S_i représente la surface de la i -ème région. Chaque région correspond à une modulation spécifique comme le montre la figure 5.1. Nous supposons que tous les utilisateurs dans une région ont un même SINR moyen et donc une même modulation et un même taux de codage.

Valeur reçue du SINR (dB)	Rayon	Modulation	Taux de codage
$[\gamma_1, \gamma_0)$	R_1	64-QAM	3/4
$[\gamma_2, \gamma_1)$	R_2	16-QAM	3/4
$[\gamma_3, \gamma_2)$	R_3	QPSK	1/2
$[\gamma_4, \gamma_3)$	R_4	BPSK	1/2

TABLE 4.1 – Partitionnement du SINR avec AMC

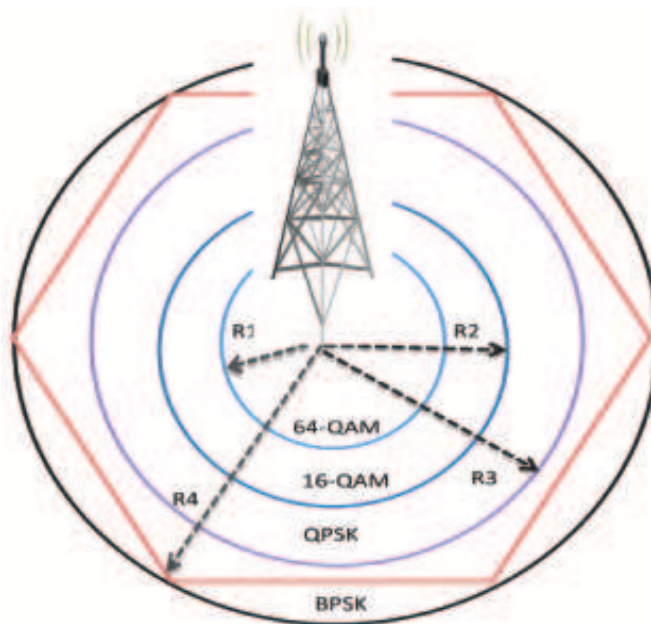


FIGURE 4.1 – La cellule HSDPA sectorisée en plusieurs régions avec AMC.

4.2.2 Bande passante par type d'appel

Nous supposons que tous les terminaux des utilisateurs supportent les modulations ci-dessus. Soit $P_{i,c}$ la puissance transmise par la station de base du canal HS-DSCH vers l'utilisateur de la région i . Soit q_i la variation des conditions du canal due à la distance entre l'utilisateur de la région i et le Node-B, aux évanouissements rapides et lentes (Bonald et Proutière, 2003; Sartori et al., 2007). L'interférence sur un utilisateur dans la région i (intra cell interférence) est donné par

$$I_{intra,i} = \alpha \frac{P_{tot} - P_{i,c}}{q_i}, \quad (4.1)$$

où P_{tot} est la puissance totale transmise par la station de base HSDPA et α est un facteur d'orthogonalité en lien descendant (Mehta et al., 2002). Soit R_c le débit binaire d'un appel de la classe- c ($c=RT, NRT$). Les utilisateurs demandant des services de la même classe dans la région i ont une même valeur moyenne du SINR dans $[\gamma_i, \gamma_{i-1})$ et un même débit R_c quelques soit la position de ces utilisateurs. La relation entre le débit, le

SINR moyen de la région i et la modulation AMC est donné par

$$\frac{R_c}{(\log_2(M)\tau)_i} = \frac{W}{SF_{16}} \frac{SINR_{i,c}}{E_c/N_0}, \quad (4.2)$$

où E_c/N_0 est l'énergie pour transmettre un bit de type ($c = RT, NRT$) (Bonald et Proutière, 2003; Viterbi, 1995); W est le débit chip; τ est le taux de codage; M est l'efficacité de modulation; SF_{16} est un facteur d'étalement (spreading factor of 16) pour le canal HS-PDSCH (*high speed-physical downlink shared channel*) et le $SINR_{i,c}$ dans la région i est donné par

$$SINR_{i,c} = \frac{P_{i,c}/q_i}{I_{intra,i} + N_0}, \quad i = 1, \dots, r. \quad (4.3)$$

Pour que le mobile reçoive tout le temps un débit constant R_c , nous cherchons à trouver une bande passante dynamique qui dépend de l'efficacité de modulation de chaque région, notée Δ_c^i . Utilisant les relations (4.1) et (4.3), nous obtenons

$$\frac{SINR_{i,c}}{1 + \alpha SINR_{i,c}} = \frac{P_{i,c}}{\alpha P_{tot} + q_i N_0}. \quad (4.4)$$

Soit M_c^i le nombre d'appels HSDPA de la class- c dans la région i . Alors la puissance totale transmise par le Node-B est

$$P_{tot} = P_{CCH} + P_{SCCH} + \sum_{i,c} M_c^i P_{i,c}, \quad (4.5)$$

où P_{CCH} est la puissance du canal commun CCH (*Common Channel*) et P_{SCCH} est la puissance associée au canal de contrôle partagé SCCH (*Shared Control Channel*). Ces deux puissances représentent une fraction de la puissance P_{tot} et donc il existe une constante positive $\psi \in (0, 1)$ telle que $P_{CCH} + P_{SCCH} = \psi P_{tot}$.

Posons $\Lambda_c^i = \frac{P_{i,c}}{\alpha P_{tot} + q_i N_0}$, déduisons la valeur de $P_{i,c}$ qui remplacée dans (4.5), donne la forme suivante

$$P_{tot} = \frac{N_0 \sum_{i=1}^r (q_i M_{RT}^i \Lambda_{RT}^i + q_i M_{NRT}^i \Lambda_{NRT}^i)}{1 - \psi - \alpha \sum_{i=1}^r (M_{RT}^i \Lambda_{RT}^i + M_{NRT}^i \Lambda_{NRT}^i)}. \quad (4.6)$$

Maintenant, la bande passante nécessaire pour maintenir un débit constant pour un utilisateur demandant un service de classe- c dans la région i est $\Delta_c^i = \alpha \Lambda_c^i$. D'où, à partir des relations (4.2) et (4.4) nous trouvons la forme

$$\Delta_c^i = \frac{R_c}{R_c + \frac{W}{SF_{16}} \frac{(\log_2(M)\tau)_i}{\alpha E_c/N_0}}, \quad i = 1, \dots, r. \quad (4.7)$$

On considère Θ_ϵ la capacité (bande passante) totale du système telle que $\Theta_\epsilon = 1 - \psi - \epsilon$, où $\epsilon > 0$ est un nombre très petit strictement positif. Alors, nous avons la contrainte suivante sur cette bande passante totale

$$\sum_{i=1}^r (M_{RT}^i \Delta_{RT}^i + M_{NRT}^i \Delta_{NRT}^i) \leq \Theta_\epsilon.$$

4.2.3 Modélisation

Nous considérons que les processus des arrivées d'appels temps réel (RT) et non temps réel (NRT) sont indépendants et arrivent selon le processus de Poisson. Ainsi que, les temps de services sont indépendants et exponentiellement distribués. Soit $\lambda_{0,c}^i$ le taux d'arrivée des nouveaux appels de classe- c dans la région i . Soit $1/\mu_{RT}$ la durée moyenne d'un appel temps réel. La durée d'appel non temps réel dépend de la taille moyenne du fichier à télécharger en bits $E(Pay)$ (Downey, 2001), elle est donnée par

$$\frac{1}{\mu_{NRT}} = \frac{E(Pay)}{R_{NRT}}.$$

Nous rappelons que la cellule est décomposée en r régions et dans chacune d'elles nous avons deux bandes passantes différentes Δ_{RT}^i et Δ_{NRT}^i . Ce qui montre qu'il y a $2r$ classes d'appels dans le système : r classes d'appels RT et r classes d'appels NRT. Nous considérons chaque région comme une file d'attente avec une infinité de serveurs. Alors, le système a r files d'attente et tous les appels d'une région restent dans une même file. Quand un utilisateur change sa file vers une autre avant qu'il ne termine son service, nous parlons de la mobilité intra cellule (entre les régions au sein de la même cellule).

4.3 Mobilité intra cellule

4.3.1 Sans mobilité

Nous considérons ici un contrôle d'admission qui gère les appels des deux classes sans la mobilité entre les régions des utilisateurs. Un nouveau appel est accepté s'il y a des ressources libres sinon il est bloqué. Les utilisateurs de chaque région restent dans celle-ci jusqu'à la fin de leur service. Dans ce cas, pas de changement de régions durant la durée de service d'un appel provenant de la région i . Ce scénario représente la non mobilité intra cellule, c'est à dire les utilisateurs localisés la première fois dans la région i avec une modulation de celle-ci, ne changent pas l'efficacité de modulation durant leur service. Soit $n_c^i(t)$ le nombre d'appels de la classe- c qui progressent dans la région i à l'instant t . L'état du système est défini par

$$\vec{n}(t) \triangleq (n_{RT}^1(t), \dots, n_{RT}^r(t), n_{NRT}^1(t), \dots, n_{NRT}^r(t)).$$

Alors, nous modélisons le processus $\{\vec{n}(t), t > 0\}$ par une chaîne de Markov de dimension $2r$. Soit $\pi(\vec{n})$ la distribution stationnaire du système à l'état \vec{n} , tel que

$$\vec{n} = (n_{RT}^1, \dots, n_{RT}^r, n_{NRT}^1, \dots, n_{NRT}^r).$$

Nous définissons un espace d'états qui représente toutes les configurations possibles en terme de nombre d'appels dans le système par

$$A_{\Theta_\epsilon} = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r (n_{RT}^i \Delta_{RT}^i + n_{NRT}^i \Delta_{NRT}^i) \leq \Theta_\epsilon \}.$$

La probabilité que le système soit dans l'état \vec{n} ($\vec{n} \in A_{\Theta_\epsilon}$) est donné en appliquant le théorème BCMP (Baskett et al., 1975; Chao et al., 2001) pour les classes multiples

$$\pi(\vec{n}) = \frac{1}{G(\Theta_\epsilon)} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i} (\rho_{NRT}^i)^{n_{NRT}^i}}{n_{RT}^i! n_{NRT}^i!}, \quad (4.8)$$

où $\rho_c^i = \lambda_{0,c}^i / \mu_c$ et $G(\Theta_\epsilon)$ est la constante de normalisation calculée par la relation

$$\sum_{\vec{n} \in A_{\Theta_\epsilon}} \pi(\vec{n}) = 1.$$

d'où

$$G(\Theta_\epsilon) = \sum_{\vec{n} \in A_{\Theta_\epsilon}} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i} (\rho_{NRT}^i)^{n_{NRT}^i}}{n_{RT}^i! n_{NRT}^i!}.$$

Un nouveau appel de class- c dans la région i demandant la bande passante Δ_c^i est bloqué s'il ne la trouve pas. Alors la probabilité de blocage de cet appel dans la région i et

$$B_c^i = 1 - \sum_{\vec{n} \in A_{\Theta_\epsilon - \Delta_c^i}} \pi(\vec{n}) = 1 - \frac{G(\Theta_\epsilon - \Delta_c^i)}{G(\Theta_\epsilon)}, \quad (4.9)$$

où $G(\Theta_\epsilon - \Delta_c^i)$ est la constante de normalisation calculée sur l'espace $A_{\Theta_\epsilon - \Delta_c^i}$.

4.3.2 Avec intra mobilité

Nous définissons le scénario de mobilité intra cellule quand les utilisateurs localisés dans une région donnée et connectés à la station de base ont la possibilité de changer cette région vers une autre. Ce changement pour le système signifie le changement d'efficacité de modulation d'un appel de classe- c en cours vers une autre efficacité. Alors, le taux d'arrivée d'appels de classe- c dans la région i est donné en distinguant deux cas :

1. un utilisateur peut se connecter au système la première fois dans la région i avec un taux $\lambda_{0,c}^i$,
2. un utilisateur localisé dans la région i peut être arrivé d'une région j vers celle-ci avec un taux $\lambda_c^{j,i}$ ($i = 1, \dots, r, j \neq i$).

Notons que $\lambda_c^{i,i} = \lambda_{0,c}^i$. Un appel en cours de communication à cause de la mobilité de l'utilisateur entre les régions (intra cellule) a son efficacité de modulation qui change. Quand l'utilisateur s'éloigne de la station de base, plus de ressources (bande passante) sont nécessaires pour qu'il maintienne son débit initial. Pour cela, le système a besoin de garder des ressources supplémentaires qui vont servir uniquement les appels en cours de mobilité. Ce type de ressource est connu également sous le nom des canaux de garde (*guard channel*) (Hong et Rappaport, 1986), et notons par Θ_m ces ressources.

En outre, la bande passante totale est $\Theta_\epsilon = \Theta_0 + \Theta_m$, où Θ_0 est une bande passante normale qui peut être utilisée par les appels en cours (en mobilité) et les nouveaux appels. Le problème est de connaître combien de ressources faut-il réserver pour gérer

les appels en mobilité sans pénaliser les nouveaux appels tout en gardons plus de ressource pour la mobilité. Cela nous amène à introduire un paramètre δ tel que $0 \leq \delta < 1$. Le choix de ce paramètre dépend largement de la politique de gestion des ressources adoptée par le fournisseur de service (*Service Provider*). Ce choix varie en fonction des périodes où la mobilité des utilisateurs est forte ou faible dans la journée. Nous obtenons alors les relations suivantes

$$\Theta_m = \delta\Theta_\epsilon \text{ et } \Theta_0 = (1 - \delta)\Theta_\epsilon.$$

Le fournisseur de service peut utiliser le paramètre δ (Θ_m) comme paramètre de contrôle pour déterminer la priorité d'appels en migration (mobilité) sur les nouveaux appels. Un nouvel appel de classe- c dans la région i est bloqué dès que la bande passante normale Θ_0 est complètement utilisée. Par contre, un appel en cours (en mobilité) est rejeté si la bande passante totale Θ_ϵ du système est occupée. D'où, nous définissons le taux d'arrivée total d'appels de classe- c dans la région i comme

$$\lambda_c^i(\vec{n}) = \begin{cases} \lambda_{0,c}^i + \sum_{j \neq i} \lambda_c^{j,i}, & \text{si } \vec{n} \in A_0; \\ \sum_{j \neq i} \lambda_c^{j,i}, & \text{si } \vec{n} \in A_\epsilon \setminus A_0, \end{cases} \quad (4.10)$$

où

$$A_0 = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r (n_{RT}^i \Delta_{RT}^i + n_{NRT}^i \Delta_{NRT}^i) \leq \Theta_0 \}, \quad (4.11)$$

$$A_\epsilon = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r (n_{RT}^i \Delta_{RT}^i + n_{NRT}^i \Delta_{NRT}^i) \leq \Theta_\epsilon \}. \quad (4.12)$$

Le temps moyen qu'un utilisateur reste dans la région i est noté $1/\mu_i^m$. Ce temps correspond au temps que passe un utilisateur en migration/mobilité dans une région avant qu'il ne termine son service. Ce temps dépend de plusieurs paramètres tels que sa vitesse, sa direction de mobilité, la forme et la taille géographique de sa région. Donc, le temps passé dans la région i est une variable aléatoire noté T_i . Pour simplifier, nous supposons qu'il est distribué selon la distribution exponentielle avec un moyen $1/\mu_i^m$. Ainsi, la durée totale d'un appel de classe- c dans le système est une variable aléatoire notée par T_c et également distribuée selon la distribution exponentielle avec une moyenne $1/\mu_c$. Notons que ces variables sont indépendantes et identiquement distribuées. Soit X_c^i une variable aléatoire représentant le temps d'activité ([Jeon et Jeong, 2001](#)) que peut passer un utilisateur de classe- c dans la région i . Cette variable est donnée par la relation $X_c^i = \min(T_c, T_i)$ dont la distribution est exponentielle de paramètre $\mu_c + \mu_i^m$. La probabilité d'état du système dans l'équilibre est donnée en utilisant la théorie des files d'attente. En nous basant sur le théorème BCMP pour les classes multiples avec la possibilité de changement de classe ([Chao et al., 2001](#)), nous obtenons l'expression suivante

$$\pi(\vec{n}) = \frac{1}{G_\epsilon} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i} (\rho_{NRT}^i)^{n_{NRT}^i}}{n_{RT}^i! n_{NRT}^i!}, \quad (4.13)$$

où $\vec{n} \in A_c$, $\rho_c^i = \frac{\lambda_c^i(\vec{n})}{\mu_c + \mu_i^m}$ et G_ϵ est la constante de normalisation donnée par

$$G_\epsilon = \sum_{\vec{n} \in A_c} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i}}{n_{RT}^i!} \frac{(\rho_{NRT}^i)^{n_{NRT}^i}}{n_{NRT}^i!}.$$

La distribution stationnaire dépend du taux d'arrivée d'appels de classe- c en mobilité dans la région i , $\lambda_c^{j,i}$, et inversement ce taux dépend du nombre d'appels dans la région i qui lui même est fonction de la distribution stationnaire.

4.3.3 Calcul des taux de migration

Le taux de migration d'appels de classe- c dans la région i à partir de la région j dépend du nombre marginal moyen d'appels dans la région j . Soit $p_{c,j}$ la probabilité qu'un utilisateur demandant un appel de classe- c dans la région j se déplace vers la région i avant qu'il termine son service. Cette probabilité est donnée par

$$p_{c,j} = P(T_j < T_c) = \int_0^\infty (1 - e^{-\mu_j^m t}) \mu_c e^{-\mu_c t} dt = \frac{\mu_j^m}{\mu_j^m + \mu_c}.$$

Le taux que les mobiles de la région j tentent de migrer due à la mobilité des utilisateurs est

$$E[n_c^j(t)] p_{c,j} \mu_j^m,$$

où $E(n_c^j(t))$ est le nombre marginal d'appels de classe- c dans la région j tel que

$$E[n_c^j(t)] = \sum_{\vec{n} \in A_c} n_c^j \pi(\vec{n}). \quad (4.14)$$

En conséquence, le taux d'arrivée d'appels en migration dans la région i provenant de la région j est

$$\lambda_c^{j,i} = \phi_{j,i} E[n_c^j(t)] p_{c,j} \mu_j^m, \quad i, j = 1, \dots, r; j \neq i, \quad (4.15)$$

où $\phi_{j,i}$ est la probabilité qu'un utilisateur dans la région j parte à cause de sa mobilité vers la région i . Cette probabilité dépend des paramètres de mobilité (vitesse du mobile, taille de la zone ou de la région, effets de dégradation de la qualité du signal, ...) telle que

$$\sum_{j=1, j \neq i}^r \phi_{j,i} = 1, \quad i = 1, \dots, r.$$

Afin de trouver le taux d'arrivée d'appels de classe- c ($\lambda_c^i(\vec{n})$) dans la région i en utilisant la relation (4.10), nous commençons d'abord par le calcul des taux de migration ($\lambda_c^{j,i}, j \neq i$) via l'algorithme 2 :

Algorithm 2 : Convergence des taux de migration

- 1: Initialiser les taux de migration dans l'équation (4.10) : $\lambda_{c,old}^{j,i} = 0$.
 - 2: Calculer la probabilité d'état d'équilibre $\pi(\vec{n})$ dans (5.21).
 - 3: Calculer le nombre moyen marginal d'appels RT et NRT dans (4.14).
 - 4: Dédire les nouvelles valeurs des taux de migration dans (4.15), notés par $\lambda_{c,new}^{j,i}$.
 - 5: Vérifier la convergence entre les nouveaux et anciens taux de migration : si $|\lambda_{c,new}^{j,i} - \lambda_{c,old}^{j,i}| < \xi$, où ξ est un nombre très petit et strictement positif, alors les nouveaux taux de migration sont utilisés dans le calcul des performances du système. Sinon, passer à l'étape 2 avec ces nouvelles valeurs comme des valeurs initiales. Les itérations sont poursuivies jusqu'à atteindre la convergence.
-

4.4 Mobilité inter cellules

L'objectif de cette partie est d'étudier l'impact des interférences inter cellules sur la bande passante effective par type d'appel et de calculer le taux d'appels en *handover*, c'est à dire les appels changeant la station de base. Nous considérons un réseau cellulaire composé de $l + 1$ cellules et supposons que celles-ci sont toutes homogènes et statistiquement identiques. Alors l'interférence dans la région i est causée par les mobiles de la cellule (intra interférence) et aussi par les stations de bases des l cellules voisines (inter interférence) tel que

$$I_{inter,i} = \sum_{j=1}^l \frac{P_{tot,j}}{q_{i,j}}, \quad (4.16)$$

où $P_{tot,j}$ est la puissance totale de transmission de la station de base j et $q_{i,j}$ est le coefficient d'atténuation du signal d'un utilisateur dans la région i et la station de base j . En conséquence, la relation (4.3) devient

$$SINR_{i,c} = \frac{P_{i,c}/q_i}{I_{intra,i} + I_{inter,i} + N_0}, \quad i = 1, \dots, r. \quad (4.17)$$

Pour plus de simplicité, nous considérons que toutes les cellules ont une même puissance totale de transmission P_{tot} . Nous utilisons le facteur d'interférence (*F-factor*) dans la région i de la cellule centrale 0 tel que (Sipila et al.)

$$F_{i,0} = q_i \sum_{j=1}^l \frac{1}{q_{i,j}}$$

Nous remplaçons $F_{i,0}$ pour toutes les régions ($i = 1, \dots, r$) par la valeur moyenne de ce facteur et notons la par F . En général, la capacité des systèmes de communications sans fil comme CDMA/HSDPA est limitée par les interférences. En effet, dans (Viterbi et al., 1994), les auteurs montrent que les interférences inter cellules augmentent avec l'augmentation du facteur *F-factor*. Pour trouver la bande passante effective en fonction de ce dernier, nous utilisons la même méthode que celle utilisée pour obtenir Δ_c^i dans (4.7).

Cette bande est

$$\nabla_c^i = \frac{(\alpha + F)R_c}{\alpha R_c + \frac{W}{5F_{16}} \frac{(\log_2(M)\tau)_i}{E_c/N_0}} = (1 + F/\alpha) \Delta_c^i, i = 1, \dots, r. \quad (4.18)$$

Maintenant, les espaces d'états A_0 et A_ϵ définis dans (4.11) et (4.12) seront obtenus en remplaçant l'ancienne bande passante Δ_c^i par la nouvelle ∇_c^i . L'état de la cellule 0 est également défini par $\vec{n} = (n_{RT}^1, \dots, n_{RT}^r, n_{NRT}^1, \dots, n_{NRT}^r) \in \mathbb{N}^{2r}$. Nous définissons le taux d'appels en *handover* comme taux d'appels de classe- c dans la dernière région de la cellule qui se déplacent vers d'autres cellules. Donc, le taux d'arrivée d'appels dans la région r avec le taux de *handover* (λ_c^{ho}) come défini dans (4.10) est

$$\lambda_c^r(\vec{n}) = \begin{cases} \lambda_{0,c}^r + \sum_{j=1}^{r-1} \lambda_c^{j,r} + \lambda_c^{ho}, & \text{si } \vec{n} \in A_0; \\ \sum_{j=1}^{r-1} \lambda_c^{j,r} + \lambda_c^{ho}, & \text{si } \vec{n} \in A_\epsilon \setminus A_0, \end{cases} \quad (4.19)$$

où $\lambda_c^{j,r}$ est le taux de migration d'appels donné par l'algorithme 2 et le taux de *handover* λ_c^{ho} est l'objet de l'algorithme 3. Puisque les cellules sont supposées toutes identiques, le nombre moyen marginal d'appels de classe- c dans la région r est le même. Il est donné par

$$E[n_c^r(t)] = \sum_{\vec{n} \in A_\epsilon} \pi(\vec{n}) n_c^r. \quad (4.20)$$

D'où, le taux d'appels en *handover* est donné par la relation

$$\lambda_c^{ho} = l\phi_l E[n_c^r(t)] p_{c,r} \mu_r^m, \quad (4.21)$$

où ϕ_l est la probabilité qu'un utilisateur dans une cellule fait un *handover* vers la cellule centrale

Algorithm 3 : Convergence des taux de handover

- 1: Initialiser le taux de *handover* dans l'équation (4.19) : $\lambda_{c,old}^{ho} = \lambda_c^{ho} = 0$.
 - 2: Calculer la probabilité d'état d'équilibre $\pi(\vec{n})$ dans (5.21).
 - 3: Calculer le nombre moyen marginal d'appels RT et NRT dans (4.20).
 - 4: Dédire la nouvelle valeur du taux de *handover* dans (4.26) : $\lambda_{c,new}^{ho}$.
 - 5: Vérifier la convergence entre la nouvelle et l'ancien valeur des taux de *handover*, si $|\lambda_{c,new}^{ho} - \lambda_{c,old}^{ho}| < \xi$, où ξ est un nombre très petit et strictement positif, alors le nouveau taux $\lambda_c^{ho} = \lambda_{c,new}^{ho}$ est utilisé avec l'algorithme 2 pour le calcul des performances du système. Sinon, passer à l'étape 2 avec ce nouveau taux comme un taux initial. Les itérations sont poursuivies jusqu'à atteindre la convergence.
-

4.5 Métriques des performances

Probabilité de blocage des nouveaux appels : Un nouvel appel de classe- c dans la région i est bloqué par le contrôle d'admission si l'état du système avec cet appel est

un état bloquant. L'état du système avant l'arrivée de cet appel est \vec{n} tel que $\vec{n} \in A_0$. Alors l'appel est bloqué dans la région i si $\vec{n} \in E_c^i$ tel que

$$E_c^i = \{ \vec{n} \in A_0 \mid \sum_{i=1}^r (n_{RT}^i \Delta_{RT}^i + n_{NRT}^i \Delta_{NRT}^i) + \Delta_c^i > \Theta_0 \}.$$

D'où, la probabilité de blocage des nouveaux appels dans la région i :

$$P_B^{i,c} = \sum_{\vec{n} \in E_c^i} \pi(\vec{n}), \quad i = 1, \dots, r. \quad (4.22)$$

Probabilité de perte d'appels en cours : La probabilité de perte d'une communication en cours est définie dans la littérature comme la probabilité de bloquer un appel en *handover* (Li et al., 2003). Cette définition est étendue vers le blocage d'appels en mobilité intra cellule (changement de modulation) dans (Elayoubi et Chahed, 2004). Dans ce travail, nous utilisons également la dernière définition. Nous rappelons que la perte d'appels en cours dépend directement de la politique de gestion de ressource adoptée par un fournisseur de service, c'est à dire le choix du paramètre δ et donc la valeur à réserver Θ_m . Un appel en migration de la région j (occupant Δ_c^j) vers la région i exige une bande passante Δ_c^i . Pour maintenir sa communication sans interruption, il faut que le système l'affecte la bande passante suivante $\Delta_c^{ij} = \Delta_c^i - \Delta_c^j$. La communication n'est jamais perdue si l'utilisateur s'approche de la station de base. Néanmoins, si l'utilisateur s'éloigne de la station de base, notre mécanisme de contrôle d'admission vérifie la disponibilité des ressources. En outre, un appel de classe- c en cours est perdu si $\vec{n} \in A_{i,j}^c$, avec

$$A_{i,j}^c = \{ \vec{n} \in A_\epsilon \mid \sum_{i=1}^r (n_{RT}^i \Delta_{RT}^i + n_{NRT}^i \Delta_{NRT}^i) + \Delta_c^{ij} > \Theta_\epsilon \}.$$

La probabilité de perte des communications en cours dans la région i est le rapport entre le taux d'appels en migration de la région $i - 1$ vers la région i perdus et le taux total d'appels qui se déplacent vers cette région

$$P_D^{i,c} = \frac{\sum_{\vec{n} \in A_{i,i-1}^c} \mu_{i-1}^m n_c^{i-1} \pi(\vec{n})}{\sum_{\vec{n} \in A_\epsilon} n_c^{i-1} (\mu_{i-1}^m + \mu_c) \pi(\vec{n})}, \quad i = 2, \dots, r. \quad (4.23)$$

Débit total et utilisation totale de la bande passante : Deux autres métriques importantes qui mesurent les performances globales du système sont le débit total et l'utilisation totale de la bande passante. Dans l'état d'équilibre du système, nous calculons le débit global dans les deux cas de mobilité (intra et inter cellules) ($\vec{n} \in A_\epsilon$, $\vec{n} \in A_{\Theta_\epsilon}$) par

$$T_{cell}^{hsdpa} = \sum_{\vec{n}} \pi(\vec{n}) \sum_{i=1}^r (n_{RT}^i R_{RT} + n_{NRT}^i R_{NRT}). \quad (4.24)$$

Egalement, la bande passante totale occupée par les appels en cours est obtenue comme la valeur moyenne des bandes passantes exigées par les appels en cours sur la bande

passante totale,

$$U^{hsdpa} = \frac{\sum_{\vec{n}} \pi(\vec{n}) \sum_{i=1}^r (n_{RT}^i Y_{RT}^i + n_{NRT}^i Y_{NRT}^i)}{\Theta_\epsilon},$$

où

$$Y_{RT}^i = \Delta_{RT}^i, \nabla_{RT}^i, \text{ et } Y_{NRT}^i = \Delta_{NRT}^i, \nabla_{NRT}^i.$$

4.6 Résultats numériques

Nous présentons ici quelques résultats numériques en nous basant uniquement sur la *path loss* comme fonction de la distance entre le mobile et le Node-B pour décomposer la surface de la cellule. La cellule centrale HSDPA est subdivisée en trois régions ($r = 3$) avec les modulations : 64-QAM 3/4 (6 bits/symbole), 16-QAM 3/4 (4 bits/symbole) et QPSK 1/2 (2 bits/symbole). L'énergie transmise d'un bit de class- c par bruit E_c/N_0 est de 3.4 dB pour les appels RT et égale à 2.7 dB pour les appels NRT. Le débit binaire constant pour les appels RT est $R_{RT} = 0.3 \text{ Mbps}$ et celui d'appels NRT est $R_{NRT} = 0.15 \text{ Mbps}$. Le débit chip W est de 3.84 Mbps, la valeur de ϵ est de 10^{-5} , la fraction des puissances des canaux SCCH et CCH est de $\psi = 0.2$. La durée moyenne d'appel RT est de 120 s, la taille moyenne des fichiers à télécharger est de $E(\text{Pay}) = 2 \text{ Mbits}$ et le facteur d'orthogonalité en lien descendant α est de 0.2. En plus, nous étudions l'impact de la faible mobilité des utilisateurs entre les régions. Les utilisateurs peuvent se déplacer entre les régions adjacentes et le temps moyen dans la région est 300 s. Ce dernier correspond à des utilisateurs se déplaçant à la vitesse 5 Km/h. Donc, l'utilisateur dans la région j choisit la région destinatrice ($i = j \pm 1$) avec la probabilité

$$\phi_{j,i} = \begin{cases} 1, & \text{si } j = 1, i = 2; \\ 1, & \text{si } j = r, i = r - 1; \\ 1/2, & \text{si } 2 \leq j \leq r - 1, i = j - 1 \text{ ou } i = j + 1; \\ 0, & \text{sinon.} \end{cases}$$

Le taux d'arrivée d'appels de classe- c dans la région i défini dans (4.10) devient

$$\lambda_c^i(\vec{n}) = \begin{cases} \lambda_{0,c}^i + \lambda_c^{i-1,i} + \lambda_c^{i+1,i}, & \text{si } \vec{n} \in A_0; \\ \lambda_c^{i-1,i} + \lambda_c^{i+1,i}, & \text{si } \vec{n} \in A_\epsilon \setminus A_0, \end{cases} \quad (4.25)$$

où $\lambda_c^{0,1} = \lambda_c^{r+1,r} = 0$ pour $i = 1, \dots, r$. La probabilité qu'un utilisateur arrive à la région r à partir des l cellules voisines ϕ_l est de $1/l$, donc le taux de *handover* est

$$\lambda_c^{ho} = l \frac{E[n_c^r(t)]}{l} p_{c,r} \mu_r^m = E[n_c^r(t)] p_{c,r} \mu_r^m. \quad (4.26)$$

Impact du scénario sans intra mobilité : Le scénario sans intra mobilité est illustré dans les figures 4.2 et 4.3. Ces deux figures présentent respectivement les probabilités

de blocage d'appels temps réel et non temps réel en fonction du taux d'arrivée d'appel temps réel pour différentes efficacités de modulation. Dans les deux figures nous remarquons que la probabilité de blocage dans la région trois où l'efficacité de modulation est QPSK est supérieure à celles des régions antérieures (c'est à dire les régions un 64-QAM et deux 16-QAM). Ceci est due au fait qu'un utilisateur dans la région trois demande plus de bande passante que dans les autres régions pour les deux trafics, c'est à dire, $\Delta_{RT}^1 < \Delta_{RT}^2 < \Delta_{RT}^3$ et $\Delta_{NRT}^1 < \Delta_{NRT}^2 < \Delta_{NRT}^3$. D'ailleurs, la différence principale entre la probabilité de blocage des nouveaux appels des classes temps réel et non temps réel est qu'il y a plus de blocage pour les appels RT. Car, les appels RT demandent dans notre environnement numérique plus de bande passante que celle demandée par les appels NRT dans la même région. Ces résultats sont dus au fait que notre stratégie d'allocation de ressources donne plus de priorité aux appels venant des régions proches de la station de base que ceux des régions de faible modulation en gardant un même débit pour tous ces appels. En outre, deux autres métriques importantes pour les performances globales du système sont le débit total et l'utilisation totale de la bande passante. Ces deux métriques sont illustrées dans la figure 4.4 en fonction du taux d'arrivée des nouveaux appels de type RT. Dès qu'un appel est accepté par le mécanisme du CAC, le système doit maintenir pour lui un débit constant durant son service. Ce qui implique que le débit total du système augmente, et aussi que l'utilisation totale de la bande passante augmente.

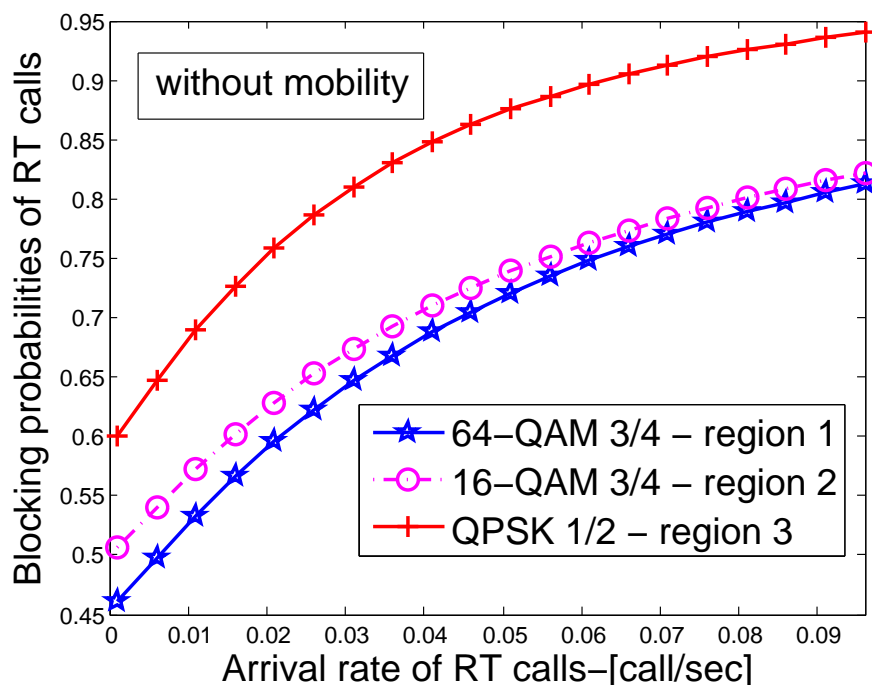


FIGURE 4.2 – Probabilités de blocage d'appels temps réel en fonction du taux d'arrivée d'appels RT pour $\lambda_{0,NRT}^i = 10^{-5}$.

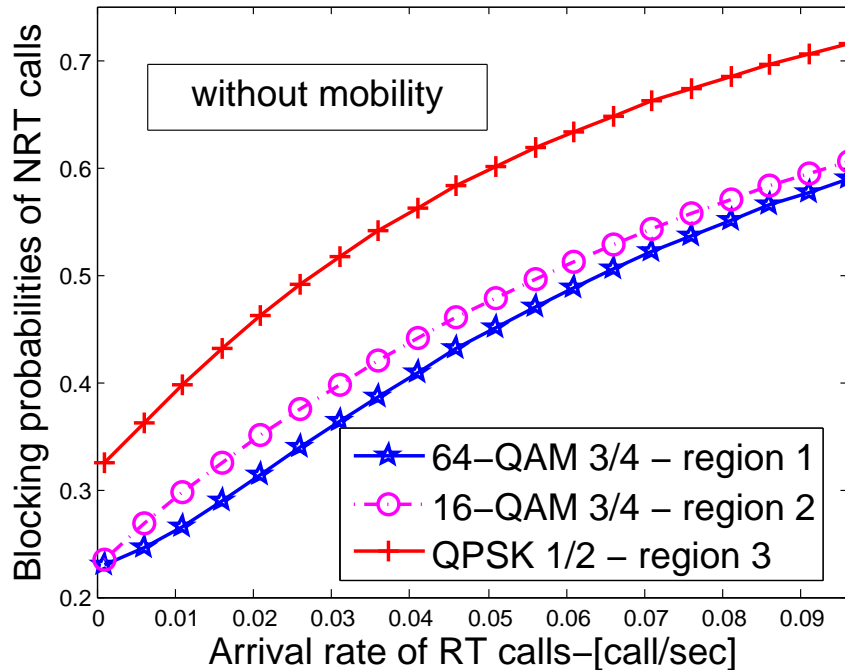


FIGURE 4.3 – Probabilités de blocage d’appels temps réel en fonction du taux d’arrivée d’appels NRT pour $\lambda_{0,NRT}^i = 10^{-5}$.

Impact du scénario de mobilité : Intéressons-nous maintenant à l’impact du scénario de mobilité intra et inter cellules sur les performances du système HSDPA. Les figures 4.5 et 4.6 illustrent la probabilité de blocage des nouvelles arrivées d’appels RT et NRT pour différentes configurations d’efficacité de modulation en fonction de la proportion (δ) de la bande passante totale réservée aux appels en mobilité. Nous remarquons que ces probabilités augmentent avec l’augmentation de cette proportion, car notre stratégie du CAC favorise par cette proportion de ressources les appels en mobilité sur les nouveaux. Ainsi, la probabilité de blocage d’appels croît avec l’augmentation de δ et donc de Θ_m . Ceci est justifié par le fait d’augmenter δ laisse moins de ressources aux nouveaux appels et donc plus d’appels d’entre eux seront bloqués par le CAC en fonction de leur efficacité de modulation. La figure 4.7 représente dans l’échelle logarithmique les probabilités de perte d’appels en cours dans le cas du scénario intra mobilité (mobilité entre les régions) en fonction des ressources réservées aux appels en mobilité. La perte d’appels en cours se produit quand ils partent de la région de haute modulation (moins de ressources par appel) vers celle de faible modulation (plus de ressource par appel). Nous constatons dans cette figure que la probabilité de perte d’appels en cours de la région 1 (64-QAM) vers la région 2 (16-QAM) est faible par rapport à celle de perte d’appels de la région 2 vers la région 3 (QPSK) et décroît jusqu’à ce que la bande passante réservée Θ_m devienne $\delta = 8\%$. Quand cette bande est supérieure à 8%, la probabilité de perte devient constante pour tous les trafics. Notons que grâce à la mobilité des utilisateurs, les appels en cours qui s’approchent de la station de base ne seront jamais bloqués par le mécanisme du CAC. En effet, ils demandent dans les

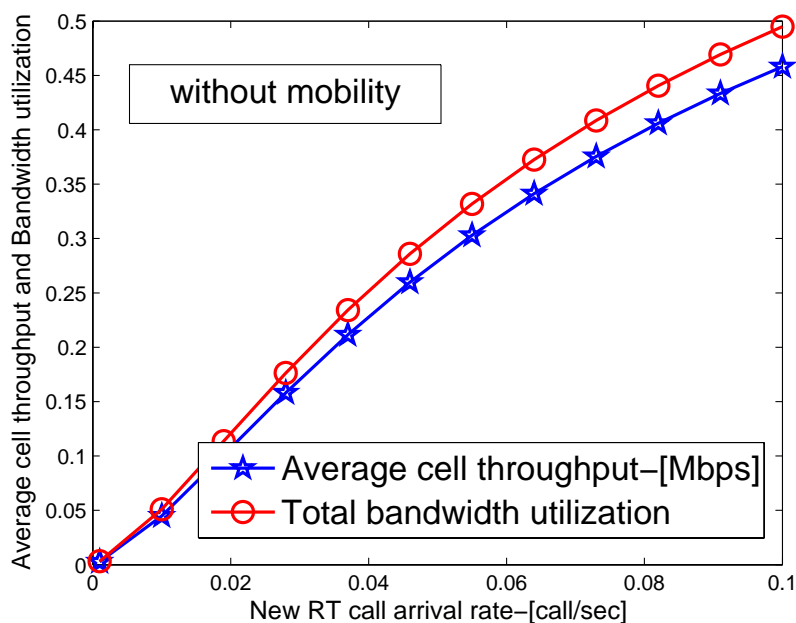


FIGURE 4.4 – Débit total et l'utilisation de la bande passante en fonction du taux d'arrivée d'appels RT pour $\lambda_{0,NRT}^i = 10^{-5}$.

régions de haute efficacité de modulation moins de ressource que dans les régions de faible efficacité de modulation ou ils exigent plus de ressource. La figure 4.8 représente le débit total en fonction de la bande passante réservée aux appels en migration avec la mobilité inter cellules. Cette figure illustre bien l'impact de l'interférence inter cellules sur les performances du système à savoir le débit total. Ce débit se dégrade avec l'augmentation du facteur d'interférence inter cellules (*F-factor*) même si on laisse plus de ressource pour la gestion d'appels en mobilité. La station de base alloue plus de ressource à un appel afin de compenser l'effet des interférences inter cellules au niveau du récepteur mobile.

Les figures 4.5 à 4.8, nous permettent de conclure que la réservation de ressource Θ_m supérieure à 8% ($\delta > 8\%$) devient inutile et ne donne pas de bonne valeurs de performance pour les probabilités de blocage et le débit total. Ainsi la probabilité de perte d'appels RT/NRT en cours est acceptable dans la région $0\% \leq \delta \leq 8\%$.

La figure 4.9 représente la probabilité de perte d'appels en cours RT et NRT en fonction de la proportion de la bande passante totale réservée avec les interférences inter cellules (*F-factor*). Ces résultats montrent que l'augmentation du facteur d'interférence implique une diminution du débit total et donc une réduction dans l'espace d'états d'appels. En outre, moins d'appels occuperont toute la bande passante du système quand on considère l'interférence inter cellules. Cette dégradation dans la capacité du système HSDPA confirme le résultat obtenu dans (Viterbi et al., 1994). D'ailleurs, la probabilité de perte d'appels en cours est élevée quand δ est dans $0\% \leq \delta \leq 8\%$ et très faible dans la région $\delta > 8\%$. En conséquence, la région $0\% \leq \delta \leq 8\%$ est un

bon compromis entre les performances du système en termes de probabilité de perte et de blocage, et le débit total. Car, avec cette région, notre mécanisme du CAC permet d'obtenir un débit total acceptable avec moins de perte d'appels en cours et moins de blocage des nouveaux appels RT et NRT. Selon ces résultats, le choix optimal de δ (Θ_m) dépend fortement du comportement du trafic que veut obtenir le fournisseur de service en fonction de sa politique de gestion des ressources entre les appels en cours et les nouveaux arrivants.

Nous supposons maintenant que le fournisseur de service réserve $\Theta_m = 8\%$ de la bande passante totale à la gestion d'appels en mobilité. En outre, la figure 4.10 représente l'utilisation totale de la bande passante occupée par les appels RT/NRT en fonction du taux d'arrivée de nouveaux appels temps réel. Dans cette figure, on remarque l'impact direct du F -factor sur la façon ou les appels utilisent la bande passante. Pour un taux d'arrivée égale 0.04 appel/s, on voit bien que 55% de la bande passante est utilisée quand l'interférence inter cellules est élevée F -factor=0.15. Par contre, 73% est occupée quand l'interférence est faible F -factor=0.05, ce qui exploite bien la capacité du système par les appels en cours. En plus, la figure 4.11 compare les probabilités de pertes pour les appels en cours RT et NRT. Les résultats sont illustrés en fonction du taux d'arrivée des nouveaux appels RT. La principale remarque est que notre mécanisme du CAC capable de garder une faible probabilité de perte pour les appels RT et NRT s'il y a moins d'interférence inter cellules.

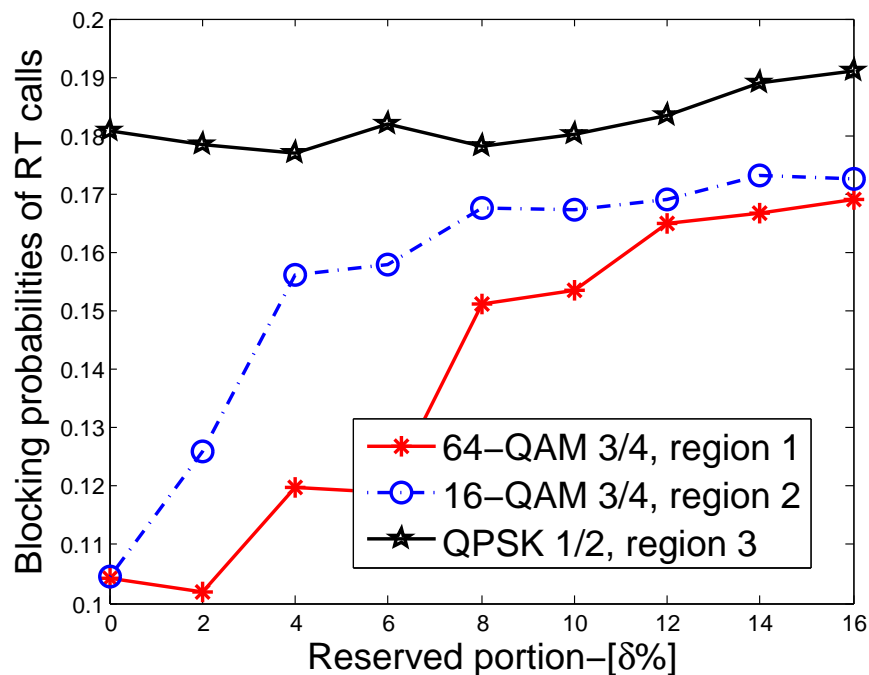


FIGURE 4.5 – Probabilités de blocage d'appels temps réel en fonction de la bande passante réservée aux appels migrant pour $\lambda_{0,RT}^i = 2 * 10^{-3}$, $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300s$.

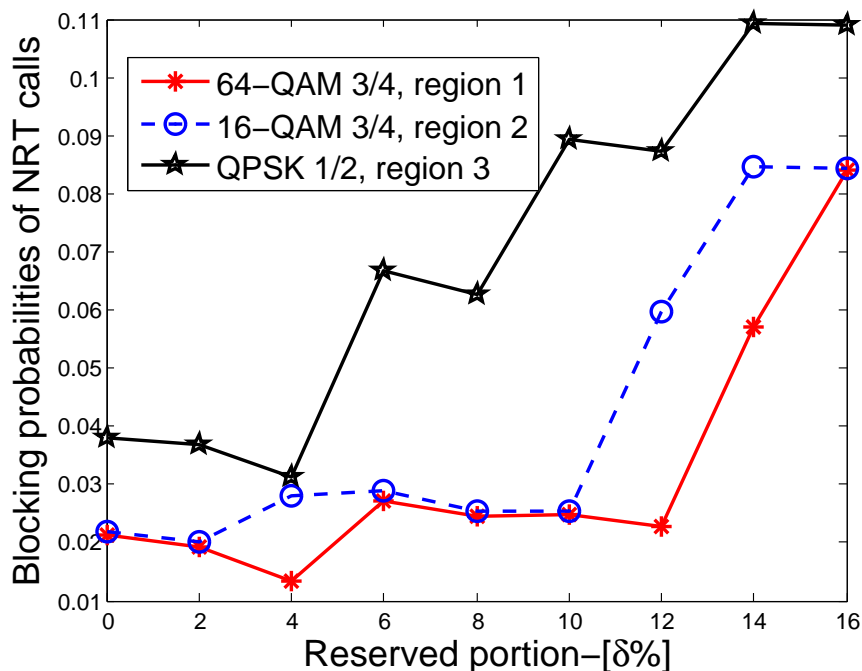


FIGURE 4.6 – Probabilités de blocage d’appels non temps réel en fonction de la bande passante réservée aux appels migrant pour $\lambda_{0,RT}^i = 2 * 10^{-3}$, $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300s$.

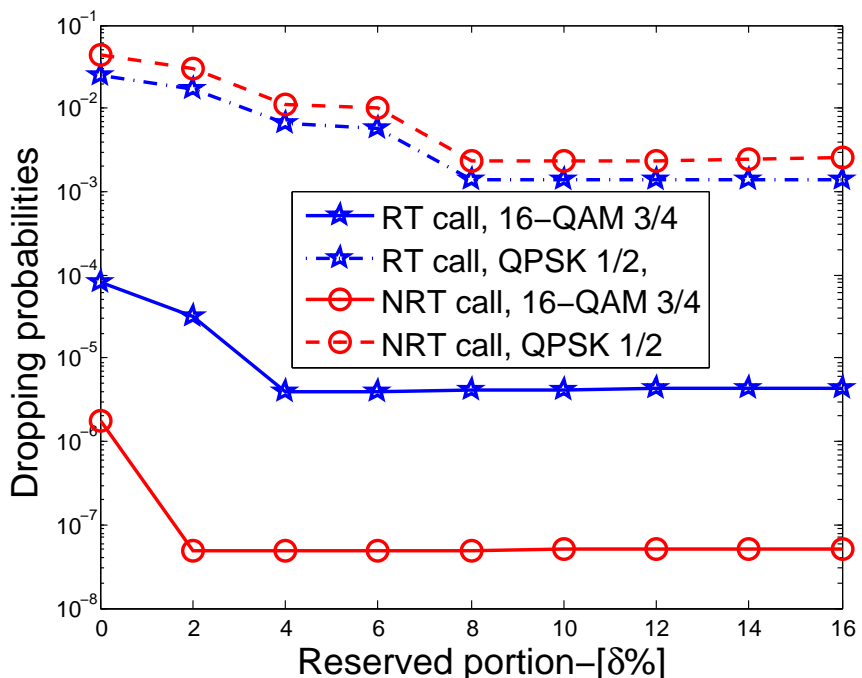


FIGURE 4.7 – Probabilités de perte d’appels temps réel et non temps réel en fonction de la bande passante réservée aux appels migrant pour $\lambda_{0,RT}^i = 2 * 10^{-3}$, $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300s$.

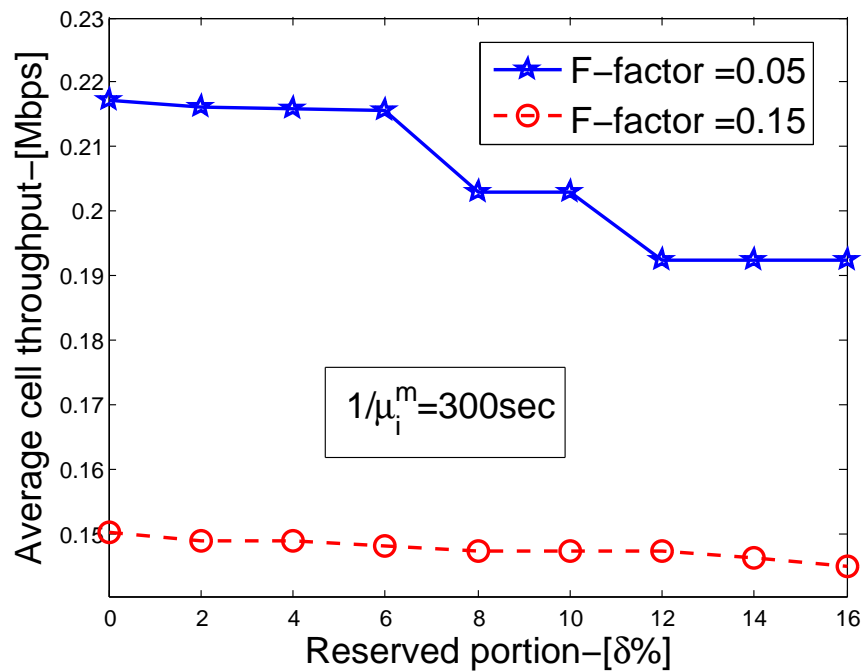


FIGURE 4.8 – Débit total en fonction de la bande passante réservée aux appels migrant pour F-factor, $\lambda_{0,RT}^i = 2 * 10^{-3}$, $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300\text{s}$.

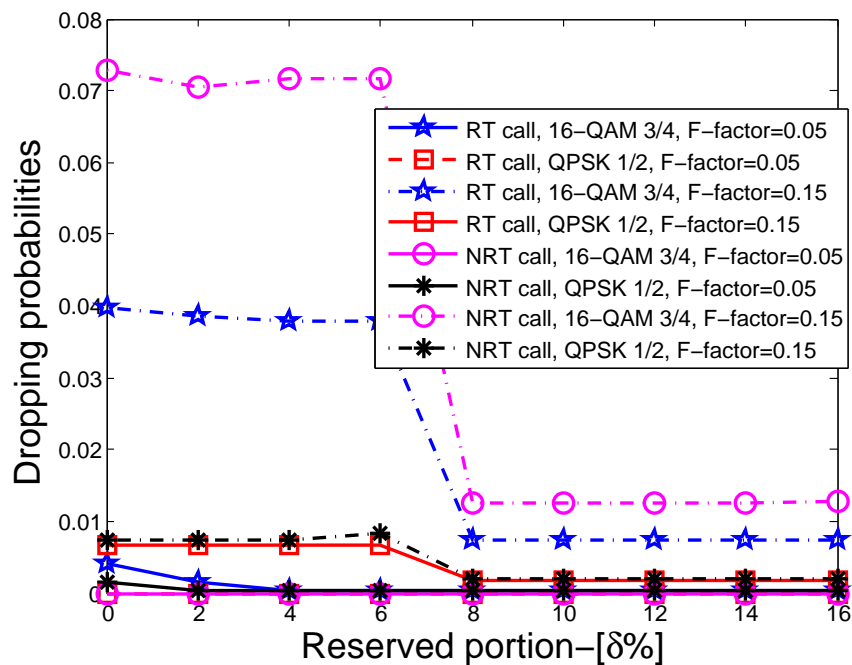


FIGURE 4.9 – Probabilité de perte en fonction de la bande passante réservée aux appels migrant pour F-factor, $\lambda_{0,RT}^i = 2 * 10^{-3}$, $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300\text{s}$.

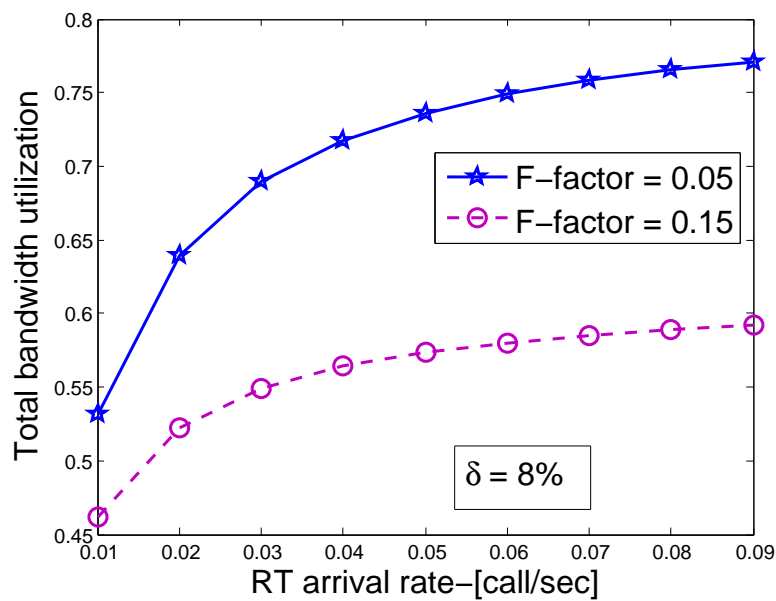


FIGURE 4.10 – Utilisation total de la bande passante en fonction du taux d'arrivée d'appels temps réel pour $\Theta_m = 0.08$ ($\delta = 8\%$), $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300s$.

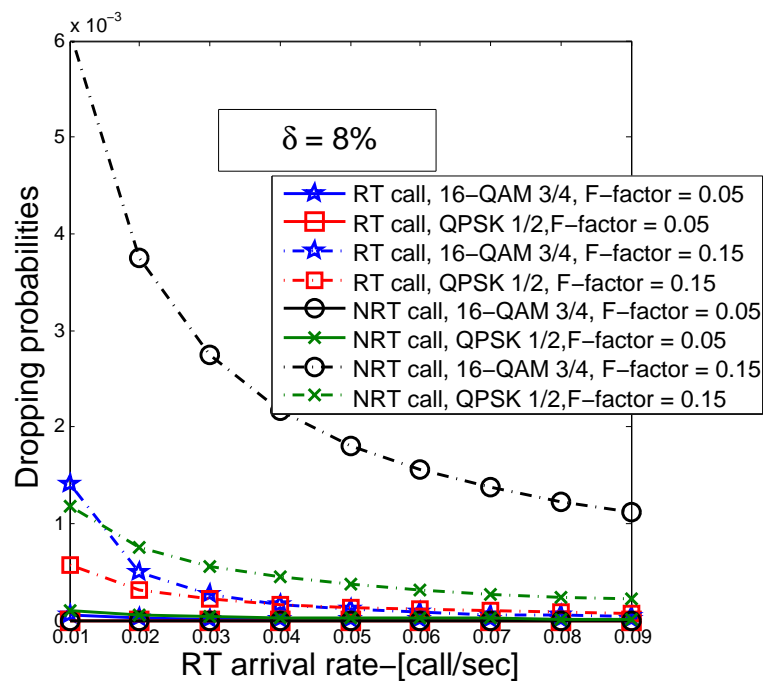


FIGURE 4.11 – Probabilité de perte en fonction du taux d'arrivée d'appels temps réel pour $\Theta_m = 0.08$ ($\delta = 8\%$), $\lambda_{0,NRT}^i = 10^{-3}$ et $1/\mu_i^m = 300s$.

4.7 Conclusion

Dans ce chapitre nous nous sommes focalisés sur l'analyse du système HSDPA comme extension du réseau 3G UMTS, et présentons une approche pour servir les utilisateurs mobiles avec un débit constant quand ils changent leur efficacité de modulation (par exemple, changeant de la haute modulation 64-QAM vers la faible QPSK et vice versa). Dans le réseau à haut débit HSDPA, l'efficacité de modulation change généralement en fonction de la qualité du signal du canal ; c'est à dire, une bonne qualité du signal permet d'affecter à l'utilisateur une haute modulation avec plusieurs bits par symbole et vice versa. Cependant, quand la modulation change de la haute valeur vers la faible, l'utilisateur a un impact négative en termes de dégradation de son débit. Ce problème est traité dans ce chapitre par l'adaptation de la bande passante quand la modulation change et dans le but de maintenir un débit constant de l'utilisateur indépendamment de sa position. Un contrôle d'admission est proposé dans ce travail opérant avec deux classes de services temps réel et non temps réel avec différents scénarios de mobilité des utilisateurs. La cellule est subdivisée en un nombre fini de régions contenant chacune une modulation donnée. Le contrôle d'admission gère les ressources du système prenant en compte la mobilité des utilisateurs entre les régions ainsi qu'entre les cellules. Le contrôle d'admission ne fait aucune priorité entre les appels des classes considérées et par contre, favorise les appels en mobilité sur les nouveaux appels en leurs les réservant certaines ressources. La réservation de ressources dépend de ce que veut faire un fournisseur de service en fonction du trafic (minimiser les probabilités de perte d'appels en cours de communication et le blocage de nouveaux appels). Le fournisseur de service peut changer la valeur réservée à la mobilité des utilisateurs selon les périodes où la mobilité est forte ou faible. Nous avons montré que la mobilité inter cellulaire et l'interférence inter cellules force notre mécanisme du CAC à allouer plus de bande passante aux mobiles pour obtenir un débit constant. Ce qui réduit l'espace d'états possibles du système et donc moins de débit total du système. Ces résultats permettent au fournisseur de service de trouver un compromis entre les performances globales du système comme le débit total et les probabilités de perte et blocage. Comme futur travail, nous pensons à développer un mécanisme du CAC qui donne uniquement la priorité aux appels temps réel alors que les appels non temps réel partageront les ressources qui restent. Ainsi, l'on de traitera différentes classes de mobilité d'utilisateurs (mobilité élevée et réduite) et l'on considèrera le cas du lien montant HSUPA (*High Speed Uplink Packet Access*).

Chapitre 5

Gestion des ressources du réseau mobile IEEE 802.16e WiMAX

Sommaire

5.1	Introduction	119
5.2	Analyse du système IEEE 802.16e dans le cas discret	121
5.2.1	Description du système	121
5.2.2	Mécanisme de contrôle d'admission	123
5.2.3	Modèle	124
5.2.4	Allocation de ressources et Matrice de transition	126
5.2.5	Métriques des performances	129
5.2.6	Exemples numériques	131
5.3	Analyse du système IEEE 802.16e dans le cas continu	137
5.3.1	Description du système	137
5.3.2	Mécanismes de contrôle d'admission	137
5.3.3	Modèle	139
5.3.4	Comportement des utilisateurs mobiles	141
5.3.5	Analyse du système avec le premier mécanisme	143
5.3.6	Analyse du système avec le deuxième mécanisme	145
5.3.7	Evaluation des performances	147
5.3.8	Applications numériques	148
5.4	Conclusion	154

5.1 Introduction

Dans ce chapitre nous nous intéressons à la gestion de ressources du réseau de quatrième génération (WiMAX, dit réseau de la 4G) basé sur le standard IEEE 802.16e. Un des objectifs fondateurs du WiMAX Forum est la recherche de l'interopérabilité : elle est obtenue par les voies de la normalisation et de la certification, et est un des enjeux majeurs du WiMAX. C'est un enjeu d'autant plus important que WiMAX est défini pour

une large bande de fréquences, de 2 à 66 GHz, dans laquelle on trouve des technologies existantes, comme le Wi-Fi, qui autorise des débits, des portées et des usages très variés. Cette multiplicité des bandes de fréquences visées, des débits, portées et usages possibles, est d'ailleurs le principal écueil qu'affronte le consommateur final. Le réseau de la 4G vise à obtenir la convergence des réseaux existants comme le Wi-Fi, UMTS, etc. Les services offerts par le WiMAX sont classés selon les paramètres définissant la qualité de service associée à une connexion. Quatre classes de services sont proposées dans le standard : le service à acceptation non-sollicitée (UGS), le service temps réel (rtPS), service de transfert de données (nrtPS) et le trafic interactif (BE). Deux grandes classes de service ont été considérées durant ce travail : temps réel (RT), correspondant aux classes UGS ou rtPS, et non temps réel ou *best effort* (NRT), correspondant aux classes nrtPS et BE. Ainsi, nous considérons le cas où les codes de partitionnement alloués aux requêtes de connexion sont distribués entre les classes d'appels. Nous considérons également dans notre étude une seule cellule WiMAX/IEEE802.16 décomposée en r régions concentriques et chacune d'elles utilise une modulation adaptative (AMC) différente, comme recommandé dans ([IEEE Std 802.16e, 2005](#)).

Le premier objectif de ce chapitre est d'utiliser les résultats de ces auteurs pour proposer un contrôle d'admission au niveau de la couche MAC IEEE802.16e. Ainsi, l'objectif du mécanisme de contrôle d'admission devra éviter aux trafics de faible priorité une famine des ressources. En nous basant sur le processus généralisé des requêtes de connexions, nous développons un nouveau mécanisme d'allocation de ressources pour les appels temps réel. Ce mécanisme se base sur la position des mobiles dans la surface couverte par la station de base. La position est obtenue en fonction de la valeur instantanée de la condition radio du mobile, dite le SNR. Dans ce mécanisme, tous les appels temps réel dans le système ont un débit binaire constant, et leur durée de transmission ne dépend ni du nombre de connexions dans le système, ni des ressources consommées mais dépend uniquement du comportement de l'utilisateur. En contre partie, la durée des appels non temps réel dépend dynamiquement des ressources qui leur sont assignées et de la position dans la cellule. Ensuite, en utilisant un modèle de chaîne de Markov en temps discret (Discrete Time Markov Chain, DTMC), nous proposons un modèle probabiliste basé sur un mécanisme de priorité multiclassés en considérant le processus d'arrivée par rafale défini dans ([Peyre et al., 2008](#)).

Le deuxième objectif de ce chapitre est de modéliser la cellule IEEE802.16e en cas continue en considérant le comportement des utilisateurs mobiles. Deux stratégies de contrôle d'admission ont été proposées et quantifiées en nous basant sur le modèle de mobilité *Random Waypoint* (RWP). Ce modèle nous permet de calculer de manière réaliste les taux de migration d'appels entre les régions de la cellule, comme indiqué par ([Hyytiä et Virtamo, 2007](#)). La première stratégie vise à garantir la même qualité de service aux appels temps réel et aux appels non temps réel. Elle limite le nombre d'appels de ces classes dans le système, minimise les probabilités de blocages et de pertes en réservant des ressources pour les appels en mobilité. La deuxième stratégie quant à elle, garde la même qualité de service des appels temps réel et introduit un partage équitable des ressources pour ceux non temps réel. Deux quantités de ressources sont réservées, l'une aux appels migrant de type temps réel, et l'autre comme une quantité

de ressources minimale pour les appels non temps réel. Aussi, ces derniers peuvent utiliser dynamiquement les ressources non utilisées par les appels temps réel. Notre analyse est basée sur la modélisation du système par une chaîne de Markov à temps continu (CMTC) pour évaluer les performances du système, en termes de probabilité de blocage et de perte, de débit et temps de séjour.

5.2 Analyse du système IEEE 802.16e dans le cas discret

5.2.1 Description du système

Nous considérons dans ce travail une seule cellule IEEE802.16e décomposée en plusieurs régions concentriques de rayons R_i et de surfaces S_i ($i = 1, \dots, r$) (voir la figure 5.1). Nous nous intéressons au lien descendant où le sens de la communication s'opère de la station de base vers le mobile. Les mobiles sont uniformément répartis sur la surface couverte par la cellule et nous supposons qu'il n'y a pas de mobilité inter et/ou intra cellulaire. La population des mobiles se répartie selon la surface de couverture des régions. Chaque région est caractérisée par une modulation choisie en fonction de la valeur du SNR reçue par la station de base d'un mobile dans cette région pour la transmission de ces données. Durant notre étude, nous considérons le cas saturé correspondant où le buffer de transmission n'est jamais vide et que les mobiles tentent toujours de transmettre des paquets selon le processus *backoff* du standard IEEE 802.16e (voir la figure 5.2). En effet, le standard IEEE802.16e offre la possibilité d'obtenir des ressources en faisant plusieurs tentatives d'envoi de requête. En premier lieu le mobile venant d'émettre une requête (*ranging request*) avec une probabilité β et un code CDMA aléatoire (entre 0 et 256) laisse le temps à la station de base de traiter et de formuler sa réponse. Ce temps d'attente est borné par un compte à rebours intitulé t_3 dans le standard. Avec une probabilité $1 - \gamma$, le mobile reçoit une réponse positive de la station de base. Au delà de son temps d'attente maximal t_r , le mobile considère que sa demande est rentrée en collision avec d'autres sur le canal de contention. Il est par ailleurs possible que la demande soit correctement arrivée à la station de base, mais que celle-ci n'a pas été retenue par le module de traitement des requêtes. Dès lors que le mobile ne reçoit pas de réponse, il engage alors un algorithme incrémental exponentiel de *backoff*. En d'autres termes, le mobile définit une fenêtre de temps pour ensuite attendre durant un temps aléatoire et uniformément réparti sur cette fenêtre. A expiration de ce temps, le mobile engage l'émission d'une nouvelle requête. A chaque échec consécutif, le mobile multiplie la fenêtre de temps du *backoff*. En outre, le standard définit un nombre maximal d'essai au bout duquel la demande de connexion est abandonnée. Dès lors qu'une tentative de connexion est satisfaite, la fenêtre de temps est réinitialisée, ainsi que le décompte des retransmissions autorisées.

Les mobiles gèrent deux classes de trafics : temps réel (RT), qui correspond aux services de la classe UGS où rtPS, et non temps réel (NRT), qui regroupe les services de la classe nrtPS où BE. Notons que le standard IEEE 802.16e définit une connexion basée sur la technique de transmission des données. Alors, pour chaque nouvelle transmis-

sion le mobile essaie une requête et en conséquent, un seul mobile peut être chargé le système avec plusieurs appels. Dans ce travail, nous utilisons la méthode des requêtes de signalisation (ranging) originalement proposée par les auteurs (Peyre et al., 2008).

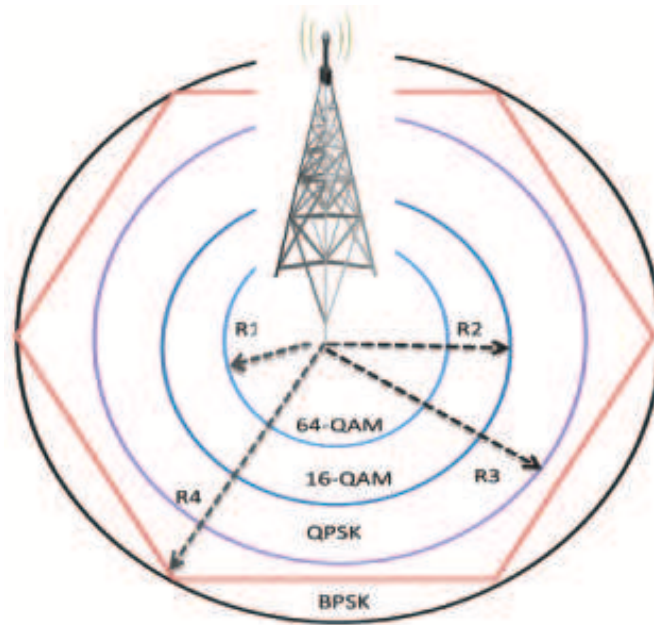


FIGURE 5.1 – La cellule IEEE802.16e sectorisée en plusieurs régions avec AMC.

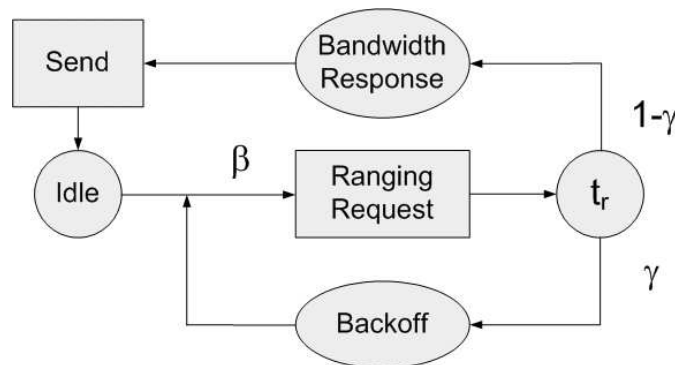


FIGURE 5.2 – Processus de connexion de backoff IEEE802.16e.

Nous considérons que la classification des codes est basée sur la requête de la bande passante, et les codes de partitionnement sont partitionnés entre les connexions des deux classes : temps réel et non temps réel. A partir du nombre de codes total CDMA, nous distinguons une sous plage de codes dédiés pour les appels temps réel et une autre sous plage partagée par les deux classes. Comme dans (Peyre et al., 2008), les arrivées des connexions temps réel et non temps réel sont dépendants et peuvent être approximer par la multiplication de deux distributions gaussiennes. En effet, La figure 5.3 présente la distribution jointe des arrivées des requêtes de connexions temps réel et non temps réel obtenues pour 50 utilisateurs exploitant 8 codes. Ces codes sont parti-

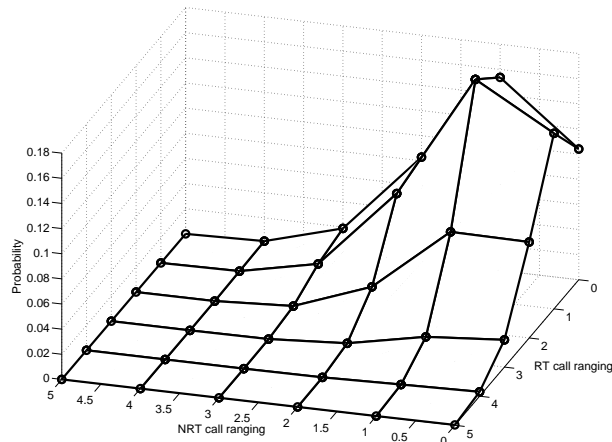


FIGURE 5.3 – Distribution des arrivées des requêtes temps réel et non temps réel, (Peyre et al., 2008).

tionnés à raison de 4 codes assignés pour les connexions temps réel et 4 autres codes partagés entre les trafics temps réel et non temps réel. Les autres paramètres suivent le standard IEEE802.16e (IEEE Std 802.16e, 2005).

5.2.2 Mécanisme de contrôle d'admission

Dans ce paragraphe nous décrivons un contrôle d'admission des requêtes arrivant au système en slots. Nous considérons un modèle à temps discret où la trame MAC est constituée d'un ensemble de slots. Les trafics temps réel sont caractérisés par un même débit binaire. En conséquence, ces appels reçoivent un nombre de porteuses correspondant à l'efficacité de leurs modulations (bit par symbole), et ainsi en fonction de leurs régions respectives. Le système peut alors accepter des communications temps réel jusqu'à débordement des ressources radio disponibles. Aussi, ces appels sont soumis à la possibilité de blocage par le système. De plus, le système a la particularité de recevoir plusieurs requêtes de ressources durant un seul temps système. Dans le cas où le système ne peut pas satisfaire toutes ces demandes à la fois, le mécanisme de contrôle d'admission acceptera un certain nombre d'entre elles en favorisant celles demandant le moins de ressources. Par conséquent, ce seront celles issues des régions les plus intérieures à la cellule qui seront satisfaites. Pour clore avec la gestion de ces trafics, nous rappelons que le temps de séjour de ce type de trafic est indépendant des ressources qu'il consomme, mais qu'il dépend uniquement du comportement de son utilisateur. Par contre, les trafics non temps réel n'ont aucune exigence en fonction de la bande passante. le CAC allouera ici un nombre de porteuses identique à chacun de ces trafics. Aussi, les trafics non temps réel gèrent la variation dynamique de leurs ressources allouées, et donc utilisent les ressources laissées disponibles par les appels temps réel déjà en cours sur la base d'un processus de partage équitable (PS) (Benameur et al., 2001).

Ainsi, le débit final d'un trafic non temps réel dépend à la fois de sa position dans la surface couverte par la station de base (c'est à dire sa région et donc sa modulation), des ressources laissées par les appels temps réel et du nombre total des appels non temps réel dans le système. Notons ici qu'afin d'éviter toute famine de ressource pour les appels non temps réel, nous intégrons dans le CAC un nombre fixe et minimum de porteuses dédiées uniquement aux appels non temps réel. Nous voulons pouvoir ainsi observer le comportement et performances de ce système en fonction de cette quantité réservée. En conséquent, les trafics non temps réel ne pourront jamais être bloqués à l'entrée du système. Enfin, les trafics temps réel restent longtemps dans le système durant un certain nombre de slots, en fonction des ressources consommées : plus le trafic non temps réel disposera de plus de sous porteuses, plus vite le trafic se terminera.

5.2.3 Modèle

Calcul théorique du débit : Nous supposons que les n mobiles sont uniformément répartis sur la surface de la couverture de la cellule. Alors le nombre n_i de mobiles qui se trouvent dans la région i est fonction de sa surface et de la population totale de la cellule. Soient r_i le rayon de celle-ci avec $r_0 = 0$ et R le rayon de la cellule. Le nombre de mobiles dans la région i ($i = 1, \dots, r$) est donc donné par la relation

$$n_i = n \frac{r_i^2 - r_{i-1}^2}{R^2}$$

A partir des relations définies dans (Tarhini et Chahed, 2007a), nous calculons le débit physique. Nous signalons que le débit (R^{RT}) d'un appel RT est fixé par les informations contenues dans la requête de demande de la bande passante à l'origine de cet appel. Nous considérons que tous les appels RT demandent au système même débit physique. Le nombre de sous porteuses assigné pour un appel RT dans la région i est alors donné en fonction de son débit demandé R^{RT} , comme suit

$$L_i^{RT} = \frac{R^{RT}}{K \times B \times E_i \times (1 - BLER_i)}, \quad (5.1)$$

où K est le nombre de fréquences assignées à chaque sous porteuses de transmission de données, B le débit en symbole (nombre de symbole par seconde), E_i l'efficacité de modulation (en nombre de bit par symbole), $BLER_i$ correspond au taux d'erreur bloc ($BLER_i$) d'une région i . Soit R_i^{NRT} le débit physique instantané atteint par les appels non temps réel dans la région i . Ce débit est calculé en fonction de la bande passante disponible aux appels non temps réel et la modulation utilisée dans la région i et donné par la relation

$$R_i^{NRT} = L_i^{NRT} \times K \times B \times E_i \times (1 - BLER_i). \quad (5.2)$$

où L_i^{NRT} est le nombre de porteuses assignées à un appel de classe NRT qui se trouve dans la région i .

Lois des arrivées et des départs des connexions : Dans cette partie, nous définissons les lois des arrivées et des départs des connexions qui se produisent dans le système. Dans le but de déterminer la capacité du système, et les performances de notre nouveau algorithme de CAC, nous commençons tout d'abord par caractériser les requêtes de ces appels. Les requêtes de types RT où NRT proviennent de chacun des r région de la cellule. Soit Z^k une variable aléatoire représentant le nombre des appels venant de la classe k , $k = \{RT, NRT\}$. Les arrivées des requêtes de signalisation de types RT et NRT sont dépendantes et peuvent approximer par la multiplication de deux distributions gaussiens. Le nombre maximum des appels RT (respectivement NRT) est fonction du nombre de codes associés à ce type de trafic. Notons que le nombre total d'arrivées ne doit pas être dépasser le nombre total de code. Nous avons dans notre cas de partitionnement des codes $Z^{RT} + Z^{NRT} \leq N_{RT}$ où N_{RT} est le nombre total de codes d'appels RT, Z^{RT} représente les codes utilisés par RT et Z^{NRT} représente ceux utilisés par NRT. Maintenant, nous supposons que toutes les requêtes des arrivées dans le système sont sans fil d'attente. Nous notons aussi par Z le processus d'arrivée : $Z = (RT_1, \dots, RT_r, NRT_1, \dots, NRT_r)$ où RT_i (respectivement NRT_i) correspond au nombre d'arrivée des appels RT (respectivement NRT) dans la région i . Nous associons la probabilité suivante au processus Z

$$\begin{aligned} P_z(a_1, \dots, a_r, b_1, \dots, b_r) &= P(RT_1 = a_1, \dots, RT_r = a_r, NRT_1 = b_1, \dots, NRT_r = b_r), \\ &= \frac{\prod_{i=1}^r \binom{n_i}{a_i} \binom{n_i - a_i}{b_i}}{\binom{n}{a} \binom{n - a}{b}} P(Z^{RT} = a, Z^{NRT} = b), \end{aligned}$$

$$\text{avec } a = \sum_{i=1}^r a_i \quad \text{et} \quad b = \sum_{i=1}^r b_i.$$

Désormais, nous désirons déterminer le processus de départ des appels du système suite à la terminaison de leurs services. Nous calculons ainsi les probabilité de départ des appels de classe k dans une région i . Sachant que la durée des appels RT ne dépend pas des ressources consommées. Les appels RT demeure dans le système durant un temps exponentiellement distribué de moyenne μ^{RT} . Cette moyenne ne varie pas au cour du temps ni en fonction de la charge globale du système.

Par contre, la durée des appels NRT dépend des ressources consommées. Aussi, le service est exponentiellement distribué de moyenne μ_i^{NRT} . Mais notons que cette moyenne change dynamiquement avec la charge du système. A chaque temps système, le nombre de service en cours évolue, ainsi que les ressources laissées libres par les appels RT. Signalons aussi que la moyenne μ_i^{NRT} dépend uniquement des ressources attribuées à chaque appel NRT en début de temps système. Cette moyenne n'est pas sujette aux libérations de ressources faites par les appels se terminant durant tout le temps système observé. En effet, le standard IEEE802.16e renseigne les mobiles de leurs ressources alloués uniquement en début de trame MAC via la carte de la sous trame du lien descendant DL-MAP. Aussi, les mobiles sont incapables de prendre connaissance des libérations de ressources éventuelles ayant lieu durant l'espace de temps d'une trame MAC

avant de recevoir les «MAP» de la trame suivante. Par conséquent, nous considérons que la bande passante disponible en début de trame MAC ne change pas durant tout l'espace de temps de cette trame.

Ainsi, le temps moyen de séjour d'un appel NRT dans la région k est fonction de la taille moyenne des paquets de données $E(Pay)$ à transmettre et du débit de cet appel R_i^{NRT} . Ce temps étant définis comme suis :

$$\frac{1}{\mu_i^{NRT}} = \frac{E(Pay)}{R_i^{NRT}}.$$

Dès lors nous pouvons déterminer le nombre moyen de fin de service durant une trame MAC. Soit n_i^k le nombre des appels dans le système pour la classe k et la région i . Soit, T la durée d'une trame MAC. Le nombre de fin de service λ_i^k pour la classe k et dans la région i est alors donné par :

$$\lambda_i^{RT} = n_i^{RT} \mu^{RT} T \quad \text{et} \quad \lambda_i^{NRT} = n_i^{NRT} \mu_i^{NRT} T. \quad (5.3)$$

Enfin, nous déterminons ici la distribution de ces fins de service en supposant qu'elles suivent une distribution de Poisson. Soit S_i^k le nombre de service terminé durant un temps système dans la région i . Soit D_i^k désigne le nombre des appels, de classe k dans la région i , terminant leur services durant la trame MAC. La distribution des fins des appels intervenant durant une trame MAC, est calculée comme suit :

$$P_d(x) = P(D_i^k = x) = \frac{P(S_i^k = x)}{\sum_{j=0}^{m_i^k} P(S_i^k = j)},$$

où

$$P(S_i^k = x) = \begin{cases} \frac{(n_i^{RT} \mu^{RT} T)^x}{x!} e^{-n_i^{RT} \mu^{RT} T}, & \text{si } k = RT, \\ \frac{(n_i^{NRT} \mu_i^{NRT} T)^x}{x!} e^{-n_i^{NRT} \mu_i^{NRT} T}, & \text{si } k = NRT. \end{cases}$$

5.2.4 Allocation de ressources et Matrice de transition

Le système gère deux classes de service sur r régions. Aussi, nous représentons un état du système en un vecteur \vec{n} . Ce vecteur est composé des appels en cours dans le système pour les deux classes pour chacune des r régions. Soit n_i^k le nombre des appels résidant dans le système pour la classe k et dans la région i ($i = 1, \dots, r$). Soient deux vecteurs représentant des appels RT et NRT du système respectivement \vec{n}^{RT} et $\vec{n}^{NRT} \in \mathbb{N}^{2r}$. Le vecteur \vec{n} est défini de la manière suivante :

$$\begin{aligned} \vec{n} &= (\vec{n}^{RT}, \vec{n}^{NRT}), \\ \vec{n} &= (n_1^{RT}, \dots, n_r^{RT}, n_1^{NRT}, \dots, n_r^{NRT}). \end{aligned}$$

Notons par L la bande passante totale du système et L_{min}^{NRT} la bande passante minimale réservée aux appels non temps réel. Les appels temps réel quant à eux, ils auront un accès prioritaire à la bande passante laissée après la partie minimale réservée :

$$L^{RT} = L - L_{min}^{NRT}.$$

Les appels RT sont assignés un certain nombre de sous-porteuses par région L_i^{RT} pris parmi L^{RT} . Tandis que, les appels NRT se partage la bande passante restante L^{NRT} équitablement selon le modèle défini dans (Benameur et al., 2001). Les relations suivantes montrent ce partage :

$$\begin{aligned} L_i^{NRT}(\vec{n}^{RT}, \vec{n}_i^{NRT}) &= \frac{L^{NRT}(\vec{n}^{RT})}{\sum_{i=1}^r n_i^{NRT}}, \\ L^{NRT}(\vec{n}^{RT}) &= L - \sum_{i=1}^r n_i^{RT} L_i^{RT}, \\ \sum_{i=1}^r n_i^{RT} L_i^{RT} &\leq L^{RT}. \end{aligned}$$

L'espace d'états E du système s'obtient en considérant tous les états possibles pour lesquels les appels RT ne dépassent pas la capacité totale du système L^{RT} :

$$E = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r n_i^{RT} L_i^{RT} \leq L^{RT} \}. \quad (5.4)$$

Maintenant, nous introduisons le vecteur \vec{n}' en tant qu'état du système au début de la trame $t + 1$. Cet état est le résultat de l'ensemble des transitions des arrivées et départs ayant eu lieu durant tout l'espace de temps de la trame t . L'état \vec{n}' est un état possible c'est-à-dire qu'il appartient à l'espace E et défini par :

$$\vec{n}' = (n_1'^{RT}, \dots, n_r'^{RT}, n_1'^{NRT}, \dots, n_r'^{NRT}) \quad .$$

En considérant toutes les transitions possibles entre les états de l'espace E . Autrement dit, les probabilités de transitions du système sont calculées entre les états \vec{n} et \vec{n}' .

Soit $x = (x_1^{RT}, \dots, x_r^{RT}, x_1^{NRT}, \dots, x_r^{NRT})$ le vecteur représentant l'évolution entre les vecteurs \vec{n} et \vec{n}' . Elle correspond au résultats de l'ensemble des arrivées et des départs intervenant durant le temps de la trame MAC t . Suivant ces définitions, nous notons les transitions d'état de la manière suivante :

$$P(\vec{n}, \vec{n}') = P(\vec{n}' = \vec{n} + x).$$

Notons que le système peut évoluer à partir de toutes les départs des appels de la classe k dans la région i . Ainsi, les arrivées de toutes les appels qui se trouvent dans la région i , $x_i^k \in [-n_i^k, n_i]$. Les probabilités de transition sont calculées en considérant toutes les transitions possibles pour chaque trafic et région. Néanmoins remarquons ici que ces évolutions dépendent des ressources disponibles pour les appels L^{RT} . En effet, le système accepte un nombre fini d'appel temps réel défini au travers l'espace

d'états défini par la relation (5.4). Selon les principes de notre mécanisme de contrôle d'admission, le système acceptera prioritairement les appels RT qui demandent moins de ressources en fonction de sous porteuse, c'est-à-dire, le système accepte les appels de la région la plus proche de la station de base. Par contre, les appels non temps réel où NRT sont acceptés tout le temps et indépendamment des appels RT en cours dans le système. L'étude nécessite alors d'introduire les cas où l'occupation en ressource des appels RT atteint les limites du système. Ce cas traite du blocage où le rejet d'un appel entrant alors que toutes les requêtes envoyées durant la trame MAC n'ont pas encore été traitées : Soit i^* ($i^* = 1, \dots, r$), la première région où au moins une requête est bloquée faute de ressources disponibles à sa satisfaction.

$$\begin{aligned} i^* &= \min \left(i \mid L^{RT} - \sum_{j=1}^r (n_j^{RT} - d_j^{RT}) L_j^{RT} - \sum_{j=1}^i a_j L_j^{RT} < 0 \right), \\ i &= 1, \dots, r. \end{aligned} \quad (5.5)$$

En conséquence, nous définissons les comportements de transition propre à ce cas en distinguant les deux cas d'études :

1. Les arrivées des appels temps réel dans le slot ne sont jamais atteindre la limite de la capacité. Donc, logiquement les états ne rejettent aucun appel de type RT. Soit, \vec{n}' un vecteur des appels est la différence entre les arrivées et les départs qui se produisent dans la région sans perte. Par contre, les appels NRT sont toujours acceptés, car ils adaptent leur bande passante occupée afin qu'un nouvel appel NRT soit accepter. Pour ce scénario, la probabilité de transition entre les états \vec{n} et \vec{n}' :

$$P(\vec{n}, \vec{n}') = \sum_{\substack{a_1=l_1 \\ \vdots \\ a_r=l_r}}^{c_1, \dots, c_r} \sum_{\substack{b_1=m_1 \\ \vdots \\ b_r=m_r}}^{e_1, \dots, e_r} P_z(a_1, \dots, a_r, b_1, \dots, b_r) D,$$

où

$$\begin{aligned} D &= \prod_{i=1}^r P_d(a_i - x_i^{RT}) P_d(b_i - x_i^{NRT}), \\ l_i &= \max(0, x_i^{RT}), \quad m_i = \max(0, x_i^{NRT}), \\ c_i &= n_i^{RT} + x_i^{RT}, \quad \text{et} \quad e_i = n_i^{NRT} + x_i^{NRT}. \end{aligned}$$

2. Les arrivées et les départs des appels RT définissent une région frontière, notée, i^* tel que ($1 \leq i^* \leq r$). Tous les appels venant de la région i pour $i = 1, \dots, i^* - 1$ sont acceptés. Par contre ceux qui viennent de la région i^* sont acceptés uniquement jusqu'à la limite $x_{i^*}^{RT}$. Ensuite les appels venants des régions $j = i^* + 1, \dots, r$ sont bloqués. Par conséquent, les probabilités de transitions sont obtenu en calculant tous les cas possibles de départ, d'arrivée et de migration durant une trame MAC. Pour cela nous définissons la fonction de Dirac δ par

$$\delta(X) = \begin{cases} 1, & \text{si } X \text{ est vrai;} \\ 0, & \text{si } X \text{ est faux.} \end{cases}$$

$$P(\vec{n}, \vec{n}') = \sum_{a_1=l_1}^{c_1, \dots, c_r} \sum_{b_1=m_1}^{e_1, \dots, e_r} \sum_{d_1=0}^{n_1^{RT}, \dots, n_r^{RT}} A \times D,$$

$$\begin{array}{ccc} \vdots & \vdots & \vdots \\ a_r=l_r & b_r=m_r & d_r=0 \end{array}$$

où

$$A = P_z(a_1, \dots, a_r, b_1, \dots, b_r),$$

$$D = \delta_1 \delta_2 \delta_3 \prod_{i=1}^r P_d(d_i) P_d(b_i - x_i^{NRT}),$$

$$\delta_1 = \prod_{j=1}^{i^*-1} \delta(n_j^{RT} + a_j - d_j = n_j'^{RT}),$$

$$\delta_2 = \delta(-x_{i^*}^{RT} \leq d_{i^*} < a_{i^*}^{RT} - x_{i^*}^{RT}),$$

$$\delta_3 = \prod_{j=i^*+1}^r \delta(d_j = x_j^{RT}),$$

où

$$l_i = \max(0, x_i^{RT}), \quad m_i = \max(0, x_i^{NRT}),$$

$$c_i = n_i^{RT} + x_i^{RT}, \quad \text{et} \quad e_i = n_i^{NRT} + x_i^{NRT}.$$

En nous basant sur toutes ces observations, nous pouvons procéder à la réalisation de la matrice de transition P du système :

$$\mathbf{P} = \left(P(\vec{n}, \vec{n}') \right), \text{ pour } (\vec{n}, \vec{n}') \in E \times E.$$

Maintenant, le vecteur de distribution stationnaire de probabilité d'états d'équilibre $\vec{\Pi}$, est donné en résolvant le système d'équations indépendantes linéaires suivant :

$$\vec{\Pi} = \{ \pi(\vec{n}) | \vec{n} \in E \}, \quad \text{avec} \quad \begin{cases} \vec{\Pi} \mathbf{P} = \vec{\Pi}, \\ \sum_{\vec{n} \in E} \pi(\vec{n}) = 1. \end{cases}$$

5.2.5 Métriques des performances

Débit total des appels non temps réel : Puisque les appels RT ont un débit fixé par les utilisateurs, nous intéressons par le calcul que du débit total moyen des appels NRT qui peut le système les accordés en fonction de leurs positions. D'où ce débit total moyen noté par R_{NRT}^{tot} est donné à l'équilibre du système par :

$$E(R_{NRT}^{tot}) = \sum_{\vec{n} \in E} \pi(\vec{n}) \sum_{i=1}^r n_i^{NRT} R_i^{NRT}.$$

Probabilité de blocage des appels temps réel : Un nouvel appel de type RT est bloqué par le système dans la région j faute de ressources disponibles lors du traitement de la requête demandant la bande passante nécessaire. La probabilité de blocage dans la région j notée par P_B^j est obtenue en calculant les différentes évolutions possible d'état du système aboutissant à ces blocages :

$$P_B^j = \sum_{\vec{n} \in E} \pi(\vec{n}) \sum_{\substack{c_1, \dots, c_r \\ a_1=0 \\ \vdots \\ a_r=0}} \sum_{\substack{e_1, \dots, e_r \\ b_1=0 \\ \vdots \\ b_r=0}} \sum_{\substack{n_1^{RT}, \dots, n_r^{RT} \\ d_1=0 \\ \vdots \\ d_r=0}} A \times D \times p_{i^*}^j, \quad (5.6)$$

avec $c_i = \min(n_i, N_{RT})$, $e_i = \min(n_i, N_{NRT})$ et

$$A = P_z(a_1, \dots, a_r, b_1, \dots, b_r),$$

$$D = \prod_{i=1}^r P_d(d_i) P_d(b_i - x_i^{NRT}),$$

La probabilité $p_{i^*}^j$ désigne qu'au moins un appel RT arrivant à la région j soit bloqué par le système, pour un i^* donné. Cette probabilité est de la forme

$$p_{i^*}^j = \begin{cases} 0, & \text{si } j < i^*; \\ 1, & \text{si } j > i^*; \\ 1 - \frac{\bar{a}_j}{a_j}, & \text{si } j = i^*, \end{cases}$$

où pour $i = 1, \dots, r$, \bar{a}_j désigne le nombre maximum des appels RT que peut accepter le système dans la région frontière et défini par

$$\bar{a}_j = \max \left(a_i L_i^{RT} - \sum_{k=1}^r (n_k^{RT} - d_k^{RT}) L_k^{RT} - \sum_{l=1}^{i^*-1} a_l L_l^{RT} - a_i L_i^{RT} > 0 \right).$$

Temps moyen de séjour des appels non temps réel : Nous rappelons que les appels temps réel ont des ressources dédiées dont lesquelles reçoivent un débit constant, ce qui montre que certains appels seront bloqués par le système. Alors que les appels non temps réel ne sont pas bloqués par le système, ils se partagent les ressources laissées disponibles par les autres types des appels. Or le temps de séjour des appels non temps réel dépend directement des ressources du système auxquelles ils auront accès. Le temps moyen prévu de transfert des appels non temps réel dans la région i est donné selon le résultat de Little :

$$T_i^{NRT} = \frac{E(NRT_i)}{\Lambda_i^{NRT}} = \frac{\sum_{\vec{n} \in E} \pi(\vec{n}) n_i^{NRT}}{\sum_{j=0}^{\min(n_i, N_{NRT})} j P(NRT_i = j)},$$

$$P(NRT_i = j) = \sum_{a_1=0}^{c_1, \dots, c_r} \sum_{b_1=0}^{e_1, \dots, e_r} P_{z_m}(a_1, \dots, a_r, b_1, \dots, b_r),$$

$$\begin{array}{c} \vdots \\ a_r=0 \\ \vdots \\ b_r=0 \end{array}$$

pour $l = 1, \dots, r$. $c_l = \min(n_l, N_{RT})$ et $e_l = \min(n_l, N_{NRT})$ avec $b_i = j$ et $e_i = j$.

5.2.6 Exemples numériques

Considérons une cellule OFDMA avec un espace FFT dont la dimension est 1024 sous-porteuses. La cellule devrait être décomposée en deux régions avec les deux types de modulation AMC : 64-QAM 3/4 ($E_2=3 \text{ bits/symbole}$) et QPSK 1/2 ($E_1=1 \text{ bits/symbole}$) respectivement. Fixons le taux d'erreur par bloque $BLER$ à zéro, la capacité maximale du système est de $L = 5$, la bande passante par appel RT dans les régions 1 et 2 respectivement est de $L_1^{RT} = 1$ et de $L_2^{RT} = 3$. Ainsi que le nombre de fréquences est de $K = 48$, la vitesse de modulation en baud est de $B = 2666 \text{ symboles/s}$ et la quantité moyenne à transmettre par les appels NRT est de $E(\text{Pay}) = 500 \text{ kbits}$ (Tarhini et Chahed, 2007b). Le débit des appels temps réel est de $R^{RT} = 384 \text{ kbps}$, la durée de la trame MAC est de $T = 0.001 \text{ ms}$ et la durée de service moyenne des appels RT est $1/\mu_{RT} = 120 \text{ s}$.

Impact de la durée des appels temps réel : Puisque le nouveau mécanisme de CAC proposé dans ce travail concerne que les appels temps réel, nous analysons l'influence de leurs durée de service sur les performances du système. En effet, la figure 5.4 représente le débit moyen des appels non temps réel en fonction de la bande passante réservée à ce type des appels. Les résultats correspondent à deux durées de service des appels temps réel : une rapide ($1/\mu_{RT} = 1$) et l'autre lente ($1/\mu_{RT} = 100$). Ces résultats montrent que le débit moyen est le même dans les deux cas. Dans les figures 5.5 et 5.6, nous observons que les variations de durée de service des appels temps réel n'ont pas d'influence notable sur les performances du système en fonction de la probabilité de blocage des appels RT et le temps de séjour des appels NRT quelque soit la position des mobiles dans la cellule ou bien quelque soit le type de modulation utilisée par la technique AMC.

Impact des codes de partitionnement : Dans ce paragraphe, nous essayons de voir l'impact du facteur de partitionnement des codes CDMA sur les performances du système. Certes, la figure 5.7 présente le débit moyen des appels non temps réel en fonction de la bande passante alloués, pour différents profils de partitionnement des codes CDMA. La première remarque est que le débit augmente linéairement par accroissement de la bande passante minimale allouée (L_{min}^{NRT}). La seconde remarque est que l'utilisation du partitionnement des code entraîne une augmentation du débit. Quand on augmente les codes de partitionnement pour les appels NRT, leur débit moyen augmente aussi, ce qui donne aux appels NRT plus de priorités. Dans ce cas, les collisions entre les requêtes des appels RT augmentent et pour les réduire, nous diminuons les

codes disponibles aux requêtes non temps réel. Par conséquent, nous augmentons l'occupation du système par les appels temps réel à cause de la réduction de ces codes.

Ensuite, la figure 5.9 représente le temps moyen de transfert de données des appels non temps réel dans les deux régions et pour différents profils de partitionnement. Comme prévu par notre analyse théorique, lorsque la bande passante minimale dédiée aux appels NRT augmente, le temps moyen de séjour de ces appels diminue fortement. D'après cette figure, nous pouvons déterminer aisément la valeur seuil de la capacité minimale où le gain obtenu par des valeurs supérieures est négligeable (temps de séjour devient négligeable). Par ailleurs, nous observons que l'utilisation du partitionnement des codes mène à diminution des performances de temps de séjour. Donc, le temps moyen de séjour des appels NRT augmente largement sur toutes les régions de la cellule.

Enfin, la figure 5.8 montre la probabilité de blocage des appels RT sur toutes les régions de la cellule pour différents profils de partitionnement. Nous observons ici que la probabilité de blocage des appels RT qui se trouvent proche de la station de base (région 1) est moins sensible à la bande passante minimale réservée aux appels NRT. Néanmoins, cette probabilité dans la région 2 où les appels RT sont loin de la station de base reste élevée pendant la variation de la capacité minimale. Cela est due au fait que notre CAC accepte en première priorité les appels RT arrivant dans les régions intérieures (les régions proches de la station de base où les appels consomment moins de ressources). Ainsi, cette figure montre aussi le fort impact négatif du partitionnement des codes CDMA sur la probabilité de blocage des appels RT. En conséquence, notre analyse pourrait être utilisée par un fournisseur de service dans le IEEE802.16e WiMAX pour qu'il détermine son propre seuil de bande passante minimale à allouer aux trafics NRT.

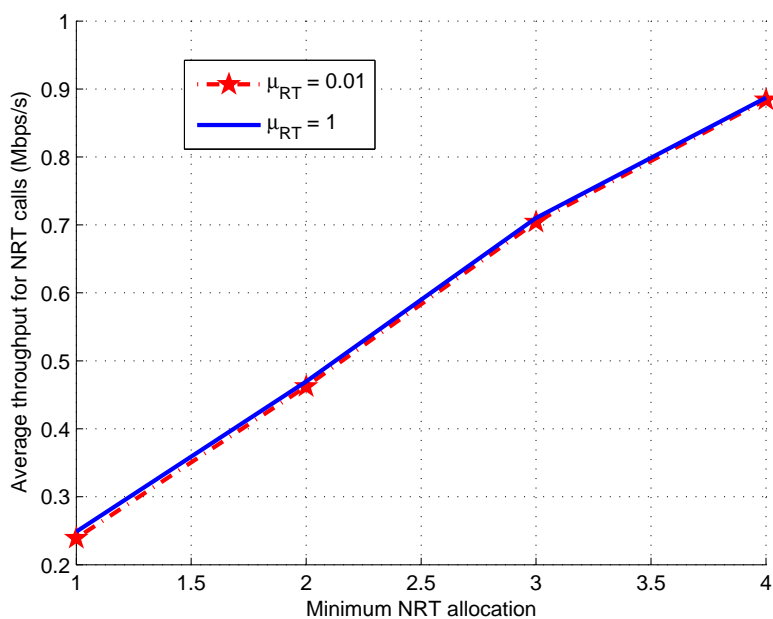


FIGURE 5.4 – Débit total moyen des appels NRT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différentes durée de service des appels RT.

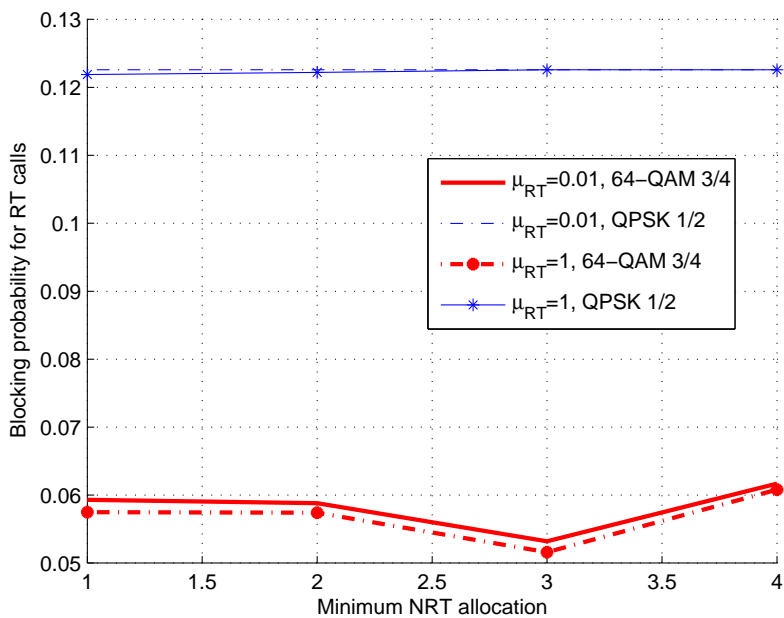


FIGURE 5.5 – Probabilité de blocage des appels RT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différentes durée de service des appels RT.

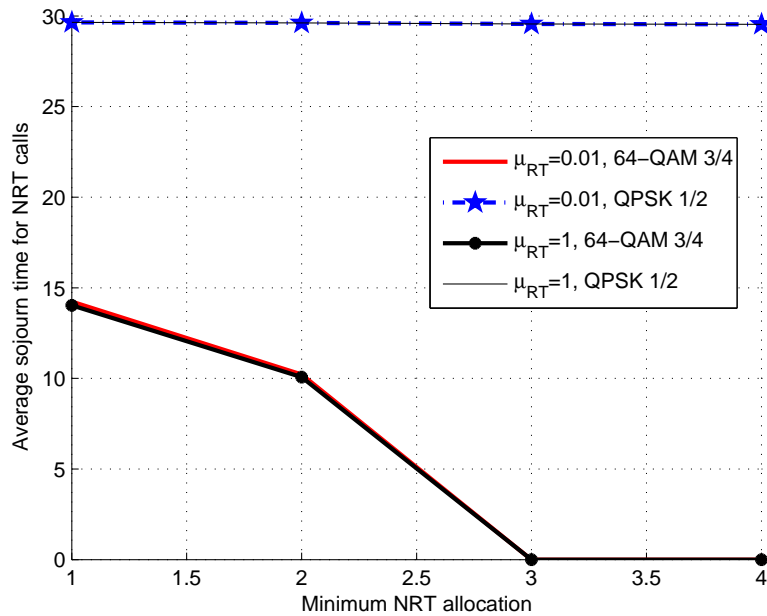


FIGURE 5.6 – Temps de séjour moyen des appels NRT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différentes durée de service des appels RT.

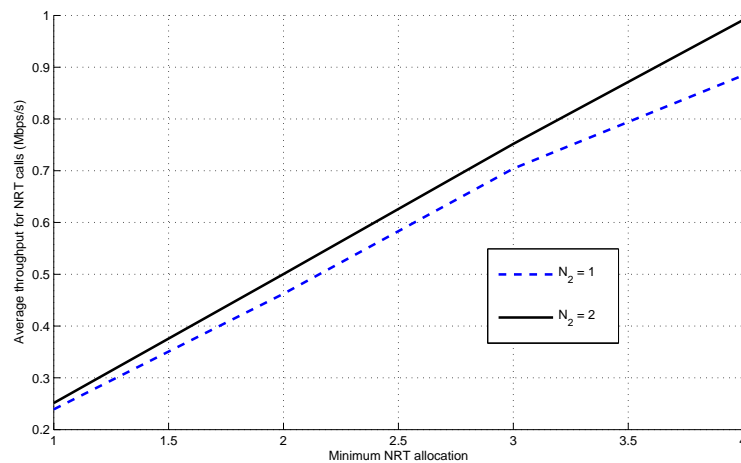


FIGURE 5.7 – Débit total moyen des appels NRT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différents codes de partitionnement.

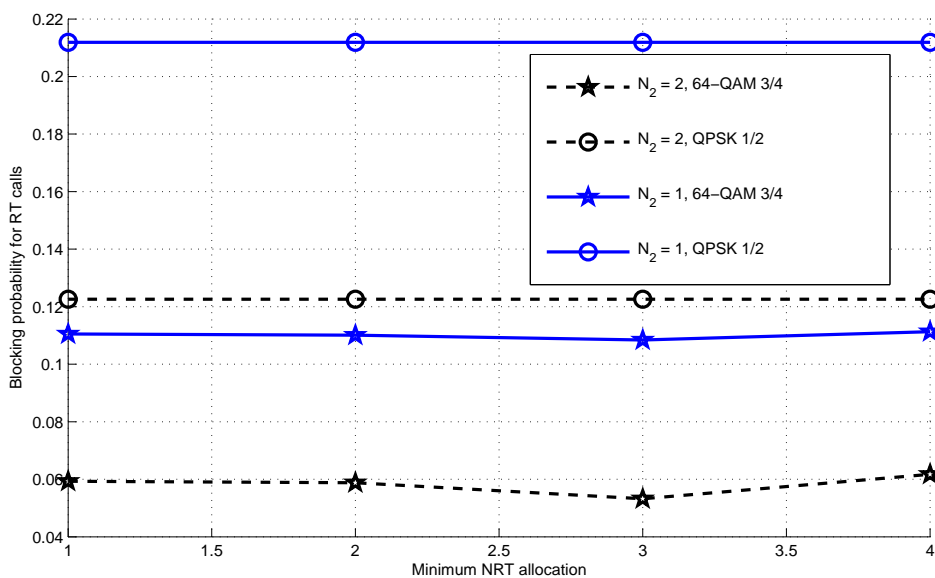


FIGURE 5.8 – Probabilité de blocage des appels RT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différents codes de partitionnement.

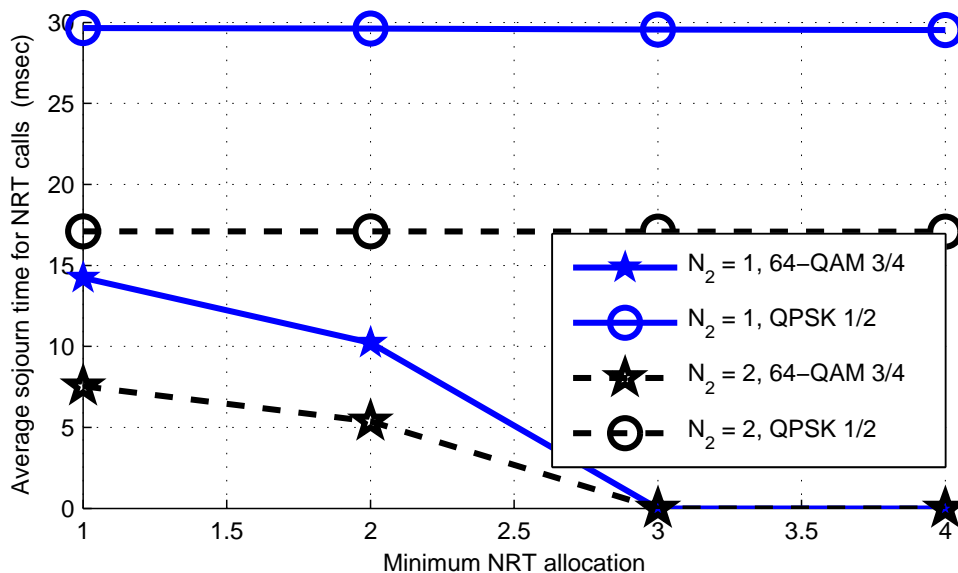


FIGURE 5.9 – Temps de séjour moyen des appels NRT en fonction du seuil minimal de la capacité réservée aux appels NRT pour différents codes de partitionnement.

5.3 Analyse du système IEEE 802.16e dans le cas continu

5.3.1 Description du système

Dans le cas discret précédent, nous avons proposé un modèle d'une cellule IEEE802.16e composée de plusieurs régions utilisant des modulations différentes. Ce modèle est basé sur un mécanisme de contrôle d'admission qui ne considère pas l'aspect mobilité des utilisateurs entre les régions. Ce qui nous a amené à proposer d'autres types de contrôle d'admission intégrant la mobilité des utilisateurs dans un système IEEE802.16e à temps continu. Egalement, la cellule est composée de r régions concentriques avec r modulations différentes. Nous considérons que les mobiles sont uniformément distribués sur l'ensemble de la cellule. Ces mobiles engagent des appels de type temps réel ou *best effort* (non temps réel). Les appels temps réel (RT) correspondant aux trafics de type UGS et rtPS, et les appels *best effort* (BE) quant à eux, correspondant aux trafics nrtPS et BE. En suite, nous modélisons le système par une chaîne de Markov à temps continu (CMTC) de dimensions $2r$. Le mécanisme de la modulation adaptative et codage (AMC) affecte une efficacité de modulation à un utilisateur selon la valeur de son SNIR.

Nous introduisons un modèle de mobilité qui s'appelle *Random Waypoint* (RWP) et qui permet de modéliser la marche aléatoire des utilisateurs dans une zone géographique convexe. Ce modèle est très célèbre en réseau ad-hoc ([Bettstetter et al., 2004](#)) et il est introduit récemment par Hyytiä et al. ([Hyytiä et Virtamo, 2007](#)) dans les réseaux cellulaires. Ce qui nous motive à le généraliser à une cellule IEEE802.16e décomposée en des régions concentriques pour déterminer les taux de migration /arrivées des appels dans chaque région de la cellule, ce qui modélise la mobilité intracellulaire. En plus, nous obtenons également les taux de *handover*.

5.3.2 Mécanismes de contrôle d'admission

Dans la figure 5.10, nous décrivons le premier mécanisme de contrôle d'admission pour un nouvel appel de classe- c .

Dans ce mécanisme, les appels reçoivent un même débit binaire selon le type de leurs classes. En conséquent, le système assigne un nombre de sous porteuses en fonction de la localisation de l'appel. L'appel est bloqué, lorsque l'allocation devient impossible due au manque de ressource. Si l'appel ne se termine pas son service dans sa première région, alors il a besoin de se déplacer vers une autre région selon la direction de sa mobilité. Cette migration implique un changement de la modulation, et ainsi, un besoin de bande passante. Les ressources disponibles du système pourraient ne pas être suffisant pour accepter une augmentation de la bande passante. Dans ce cas, la migration d'appel est perdue. La figure 5.11 quant à elle, représente notre dernier mécanisme de contrôle d'admission uniquement pour un nouvel appel de type temps réel.

Dans ce dernier mécanisme, les appels RT sont caractérisés par un même débit binaire. Ainsi, un appel de type RT reçoit un nombre de sous porteuses en fonction de son

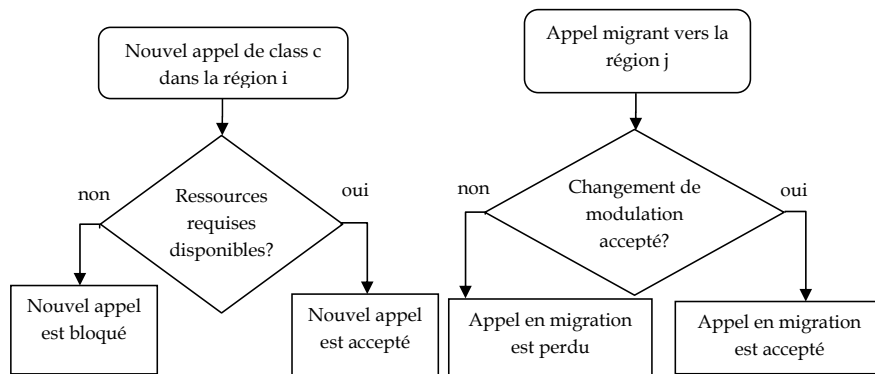


FIGURE 5.10 – Mécanisme 1 : Stratégie de contrôle d’admission d’une cellule IEEE802.16e.

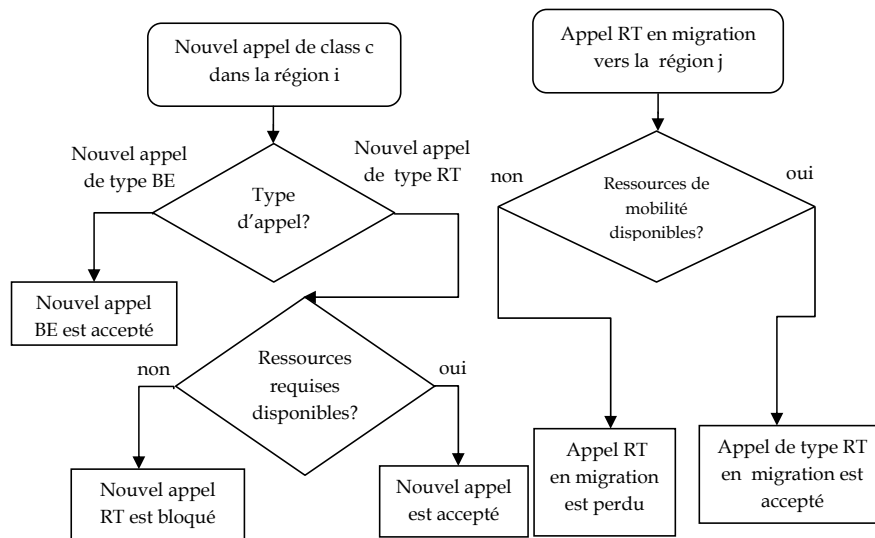


FIGURE 5.11 – Mécanisme 2 : Stratégie de contrôle d’admission d’appel temps réel d’une cellule IEEE802.16e.

efficacité de modulation et de la région ciblée. Le système accepte les appels RT jusqu’à le dépassement de la capacité. Ainsi, ils pourraient être bloqués par le système. Par contre, les appels de type BE reçoivent un même nombre de sous porteuses. Étant donné que les appels BE tolèrent la réduction de débit, ils utiliseront les sous porteuses laissées par les appels RT en cours de manière équitable selon *Processor Sharing* (PS) (Benameur et al., 2001). Aussi, le débit d’un appel BE dépend de sa région (c’est-à-dire la modulation), mais aussi du nombre d’appels BE en cours. Grâce à cette politique de gestion des ressources pour les appels BE, le système ne va jamais bloquer un appel de ce type.

Enfin, la durée d’appels RT est indépendant des ressources consommées, et donc, le temps qui va rester un appel RT dans le système dépend uniquement du comportement de l’utilisateur. Inversement, les appels BE demeurent dans le système en fonction de

la consommation des ressources, et donc, plus le nombre de sous porteuses augmente, plus vite l'appel BE termine son service.

En outre, nos mécanismes du CAC cherchent à réduire la probabilité de perte : la probabilité qu'un appel en cours de communication soit rejeté à cause de sa mobilité qu'elle change la région. Comme expliqué ci-dessus, l'appel consomme la bande passante en fonction de la modulation utilisée. Par la migration à une région loin, un appel peut exiger une bande passante additionnelle. Ainsi, cet appel pourrait subir une perte due au manque de ressources disponibles. Pour empêcher cette perte, nos mécanismes du CAC introduisent une partie de la bande passante réservée aux appels en situation de mobilité. Cette réservation vise à satisfaire au besoin des ressources additionnelles exigées en cas de migration vers les régions où la modulation est plus élevée.

5.3.3 Modèle

Débit instantané et sectorisation de cellule IEEE802.16e : Nous considérons durant notre analyse et sans perte de la généralité, la modulation adaptative et du codage (AMC) avec uniquement le *pathloss*. En fonction de ce dernier, la cellule IEEE802.16e OFDMA est décomposée en des r concentriques régions. Chaque région correspondant à une certaine intervalle de valeurs de SNR, et donc, à une AMC, comme montré dans la figure 5.12 et le tableau 5.1. Soient R_i ($i = 1, \dots, r$) le rayon de la région i et S_i représente la surface correspondante. Dans l'environnement OFDMA, le nombre total N de porteuses est divisé en L sous porteuses ou groupe contenant chacun k fréquences, tel que $k = N/L$. Dans notre étude, également, nous considérons le système multi-services IEEE802.16e WiMAX avec deux types de trafics temps réel (RT) et best-effort (BE). En plus, nous définissons le débit binaire instantané ou taux de transmission sur l'interface radio d_c^i pour un appel de class- c situé dans la région i comme suit :

$$d_c^i = L_c^i \times k \times B \times e_i, \quad (5.7)$$

où k est le nombre de fréquences assignées à chaque sous porteuses de transmission de données, B le débit en symbole (nombre de symbole par seconde), e_i l'efficacité de modulation (en nombre de bit par symbole) et L_c^i est le nombre de porteuses assignées pour l'appel de classe- c dans la région i . Le débit binaire ci-dessus peut être dégradé par les erreurs du canal dû à la collision, l'effet d'ombre, etc, comme défini dans (Tarhini et Chahed, 2007b),

$$R_c^i = d_c^i \times (1 - BLER_i), \quad (5.8)$$

où $BLER_i$ correspond au taux d'erreur bloc (BLER) d'une région i . Les modulations et les codages utilisés dans une cellule IEEE802.16e en fonction de la valeur du SNR d'un utilisateur sont résumés dans le tableau 5.1. Les valeurs du SNR dans ce tableau sont données pour un BLER inférieur à 10^{-6} dépendent du type de modulation comme indiqué par le standard (IEEE Std 802.16-2001, 2002). Et ainsi, nous avons $\gamma_1 = 24.4$ dB, $\gamma_2 = 18.2$ dB, $\gamma_3 = 9.4$ dB, $\gamma_4 = 6.4$ dB et $\gamma_0 = \infty$.

Modulation	Taux de codage	Valeur du SNR (dB)	Surface (%)
64-QAM	3/4	$[\gamma_1, \gamma_0)$	1.74
16-QAM	3/4	$[\gamma_2, \gamma_1)$	5.14
QPSK	1/2	$[\gamma_3, \gamma_2)$	20.75
BPSK	1/2	$[\gamma_4, \gamma_3)$	39.4

TABLE 5.1 – AMC selon IEEE802.16e.

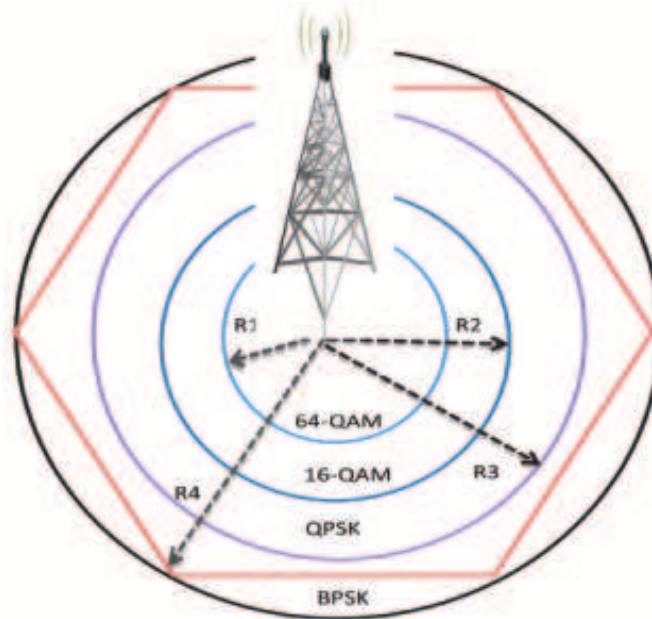


FIGURE 5.12 – La cellule OFDMA IEEE802.16e séctorisée.

Transitions et état d'équilibre : Nous nous basons sur la technique de la chaîne de Markov à temps continue (CTMC) dans notre modèle. Les différents taux de transition dans l'espace d'états réalisables définis dans les prochaines sections, sont provoqués par un des événements suivants : un nouvel appel de classe- c arrivant dans la région i ; migration d'un appel en cours de communication dans la région i vers la région j ; fin de service d'un appel de classe- c dans la région i . Les appels débutent et prennent fin indépendamment. En plus, nous prenons en considération durant notre analyse les hypothèses :

1. Le processus d'arrivée de nouveaux appels de classe- c dans la région i suit la loi de Poisson de paramètre $\lambda_{c,i}^0$;
2. Le temps de service d'un appel de classe- c dans la région i est exponentiellement distribué avec un moyen $1/\mu_c$;
3. Le temps moyen de pause ou de séjour d'un appel de classe- c dans la région i est exponentiellement distribué avec un moyen $1/\Gamma_c^i$;
4. Le taux d'arrivée moyen d'appel de classe- c migrant de la région i vers la région j est $\lambda_c^{i,j}$.

Soit $n_c^i(t)$ le nombre d'appels de classe- c à l'instant t dans la région i . Nous définissons l'état du système à l'instant t par : $\vec{n}(t) \triangleq (n_{RT}^1(t), \dots, n_{RT}^r(t), n_{BE}^1(t), \dots, n_{BE}^r(t))$. Ensuite, nous modélisons le processus $\{\vec{n}(t), t > 0\}$ par une chaîne de Markov qui correspond au processus quasi de naissance et de mort. La distribution d'états d'équilibre du système est unique et nous avons :

$$\vec{n} = (n_{RT}^1, \dots, n_{RT}^r, n_{BE}^1, \dots, n_{BE}^r).$$

5.3.4 Comportement des utilisateurs mobiles

Dans cette partie nous introduisons le modèle de mobilité *Random Waypoint* (RWP) afin de calculer les taux réels de migration. Dans le modèle RWP un noeud se déplace dans un domaine convexe noté $\Omega \subset \mathbb{R}^2$ le long d'un segment de ligne droite entre un point (*waypoint*) et un autre. Notons ce point par P_i et sont uniformément distribués dans le domaine Ω tel que $P_i \sim U(\Omega)$. La transition d'un point P_{i-1} vers P_i est désignée comme le i^{eme} leg, et la vitesse du noeud sur le i^{eme} leg est une variable notée v_i ; $v_i \sim v$. En particulier, dans le modèle RWP les variables aléatoires v_i et P_i sont indépendants, ainsi v_i sont uniformément distribués. En respectant ces notations du modèle RWP, le processus pour un noeud est défini comme une séquence infinie des triples (Bettstetter et al., 2004),

$$\{(P_0, P_1, v_1), (P_1, P_2, v_2), \dots\}.$$

Nous considérons le domaine convexe Ω durant notre étude comme un disque et leg comme un chemin entre deux points dans ce disque. Aussi, le mobile ou l'utilisateur correspond au noeud dans RWP. Or, le processus RWP est à temps réversible *time reversible*. Ceci signifie que les taux d'arrivée à travers n'importe quelle ligne segment ou une frontière sont égaux dans toutes les directions. Alors, le taux moyen des appels qui se déplacent de la région i vers la région j par unité de temps est égale celui des appels migrants de la région j vers la région i par unité de temps, c'est à dire, $\lambda_c^{i,j} = \lambda_c^{j,i}$. Ceci est prouvé par la théorie des chaînes de Markov dans (). Notre objectif dans cette partie est de calculer les taux de migration $\lambda_c^{i,j}$. La vitesse de l'utilisateur ou du mobile est supposée uniformément distribution entre v_{min} ($v_{min} > 0$) et v_{max} . La fonction de densité de probabilité $f_v(v)$ est donnée par

$$f_v(v) = \begin{cases} \frac{1}{v_{max} - v_{min}}, & \text{si } v \in [v_{min}, v_{max}]; \\ 0, & \text{sinon.} \end{cases}$$

Soit T_i le temps de transition sur le i^{eme} leg (chemin) et il est défini par $T_i = \frac{l_i}{v_i}$, où $l_i = |P_i - P_{i-1}|$. Les variables l_i et v_i sont indépendants et le temps moyen entre un point et un autre sur le même chemin est donné par :

$$E[T] = \bar{l} \int_v \frac{1}{v} f_v(v) dv = \bar{l} \frac{\ln(v_{max}/v_{min})}{v_{max} - v_{min}} = \bar{l}.E[1/v]. \quad (5.9)$$

Notre objectif est de calculer le taux d'arrivée dans la cellule de rayon R_{i-1} . Nous considérons ici la surface A_i de la cellule de rayon R_i comme un disque convexe dans lequel les utilisateurs se déplacent selon le modèle de mobilité RWP. Nous suivons les mêmes définitions des deux distances a_1 et a_2 introduites dans (Hyytiä et Virtamo, 2007). La distance $a_1 = a_1(R_{i-1}, \phi)$ indique la distance entre le point $R_{i-1} = \|(0, R_{i-1})\| \in A_i$ et la frontière du disque A_i dans la direction de l'angle ϕ (angle antihoraire entre la tangente au point $R_{i-1} = \|(0, R_{i-1})\| \in A_i$ et a_1). La distance $a_2 = a_1(R_{i-1}, \phi + \pi)$ représente la distance dans le sens opposé de a_1 (voir la figure 5.13). Notons par $\psi(R_{i-1}, \phi)$ le spécifique flux au point R_{i-1} dans la direction ϕ et donné par

$$\psi(R_{i-1}, \phi) = \frac{1}{2C_v} \cdot a_1 a_2 (a_1 + a_2), \quad (5.10)$$

où $C_v = \bar{l} A_i^2 \cdot E[1/v]$. Nous signalons que la surface du disque de rayon R_i est $A_i = \pi R_i^2$ et la longueur moyen d'un chemin dans ce disque est

$$\bar{l} = \frac{1}{A_i} \int_0^\pi a_1 a_2 (a_1 + a_2) d\phi. \quad (5.11)$$

Le modèle de mobilité RWP appliqué dans le contexte des réseaux cellulaires sans fil par (Hyytiä et Virtamo, 2007), donne le taux d'arrivée d'un utilisateur mobile dans la région $i - 1$ sur tout le contour du disque de rayon R_{i-1} ,

$$\lambda(R_{i-1}) = 2\pi R_{i-1} \int_0^\pi \sin(\phi) \psi(R_{i-1}, \phi) d\phi. \quad (5.12)$$

A partir de la figure 5.12, nous déduisons les distances a_1 et a_2

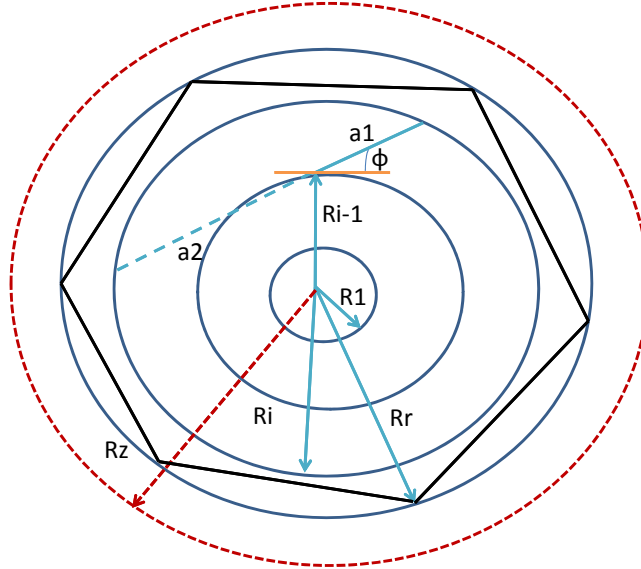


FIGURE 5.13 – Le domaine convexe du processus RWP de rayon $R_z = 2R_r - R_{r-1}$.

$$a_1(R_{i-1}, \phi) = \sqrt{R_i^2 - R_{i-1}^2 \cos^2(\phi)} - R_{i-1} \sin(\phi),$$

$$a_2(R_{i-1}, \phi) = \sqrt{R_i^2 - R_{i-1}^2 \cos^2(\phi)} + R_{i-1} \cos(\phi).$$

Puisque les comportements de mobilité des utilisateurs sont indépendants, le taux total de migration ou d'entrer dans la région j ($j = i \pm 1$) à partir de la région i ($i = 1, \dots, r$) de classe- c est

$$\lambda_c^{i,j} = \lambda(R_j).n_c^i. \quad (5.13)$$

Enfin, le taux d'arrivée d'appels venant d'autres cellules voisines (*handover*) est donné par

$$\lambda_c^{ho} = \lambda_c^{r+1,r} = \lambda(R_r).n_c^r. \quad (5.14)$$

Après avoir trouver les taux effectifs de migration d'appels, nous calculons le temps moyen de séjour d'un mobile ou d'un utilisateur dans la région i qui demande un appel de classe- c . L'existence de ce mobile dans la région i peut être qu'il vient de se connecter pour la première fois avec un taux d'arrivée $\lambda_{c,i}^0$ ou bien il vient de la région voisine j avec un taux d'arrivée $\lambda_c^{j,i}$ ($i = j \pm 1$). Alors le temps de séjour moyen est donné dans (Hyttiä et Virtamo, 2007) par la relation

$$\Gamma_c^i = \frac{p_c^i}{\lambda_c^{i+1,i} + \lambda_c^{i-1,i} + \lambda_{c,i}^0}, \quad \text{avec } \lambda_c^{0,1} = 0, \quad (5.15)$$

où p_c^i est la probabilité de trouver un appel de classe- c dans la région i , donnée par

$$p_c^i = P(n_c^i(t) \geq 1) = \sum_{\vec{n}, n_c^i \geq 1} \pi_k(\vec{n}). \quad (5.16)$$

où π_k est la probabilité d'état d'équilibre qui va être calculer, premièrement dans le cas du premier CAC ($k = 1$) et deuxièmement dans le cas du deuxième CAC ($k = 2$).

5.3.5 Analyse du système avec le premier mécanisme

Calcul de la bande passante occupée et l'espace d'états : Dans cette partie tous les appels acceptés par le système vont recevoir un même débit binaire partout dans la cellule. le mécanisme de contrôle d'admission décrit dans la figure 5.10, assure ce type de la qualité de service en fonction de la mobilité inter et intra cellulaire. Le système assigne via l'équation (5.8) à chaque appel de classe- c la quantité

$$L_c^i = \frac{R_c}{k \times B \times e_i \times (1 - BLER_i)}. \quad (5.17)$$

La durée d'un appel temps réel est déterminée en fonction du comportement de l'utilisateur indépendamment des ressources consommées. Alors que celle d'un appel *best effort* est donnée en fonction de la quantité moyenne en bits à transmettre ($E(Pay)$ (Downey, 2001)) par l'utilisateur et les ressources assignées, comme suit

$$\frac{1}{\mu_{BE}} = \frac{E(Pay)}{R_{BE}}.$$

Notre première stratégie de CAC favorise les appels en mobilité sur ceux qui tentent de se connecter la première fois dans le système. Pour cela, notons par L_m la capacité réservée strictement aux appels qui vont être déplacer à cause de la mobilité des utilisateurs.

Aussi, soit L_0 la capacité restante pour tous les appels et donnée par $L_0 = L - L_m$. La bande passante occupée à l'état \vec{n} est notée par $B_1(\vec{n})$ et définie par la relation

$$B_1(\vec{n}) \triangleq \sum_{i=1}^r \{n_{RT}^i L_{RT}^i + n_{BE}^i L_{BE}^i\}. \quad (5.18)$$

Nous définissons l'espace d'états possibles de toutes les configurations possibles de nombre d'appels dans le système par la relation

$$\mathbb{E} = \{\vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r \{n_{RT}^i L_{RT}^i + n_{BE}^i L_{BE}^i\} \leq L\}. \quad (5.19)$$

Distribution stationnaire : Nous rappelons qu'un appel de la classe- c dans la région i de la cellule peut être venir de l'extérieur du système où à partir des régions voisines sous l'effet de la mobilité. Donc, pour un état \vec{n} possible, le taux total d'arrivée d'appels de classe- c dans la région i est

$$\lambda_c^i(\vec{n}) = \begin{cases} \lambda_{c,i}^0 + \lambda_c^{i-1,i} + \lambda_c^{i+1,i}, & \text{si } B_1(\vec{n}) < L_0; \\ \lambda_c^{i-1,i} + \lambda_c^{i+1,i}, & \text{si } L_0 \leq B_1(\vec{n}) < L. \end{cases} \quad (5.20)$$

Puisque nous avons deux classes de service RT et BE, ainsi chaque appel de chacune de ces classes demande dans la région i une quantité de sous porteuses différente L_c^i , alors le système considère $2r$ classes différentes. En effet, nous avons $2r$ L_c^i valeurs différentes et donc par application du théorème multiclass de BCMP (Chao et al., 2001) avec la possibilité de changement de classe (de modulation AMC), nous trouvons la distribution stationnaire du système suivant

$$\pi_1(\vec{n}) = \frac{1}{G} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i}}{n_{RT}^i!} \frac{(\rho_{BE}^i)^{n_{BE}^i}}{n_{BE}^i!}, \quad (5.21)$$

où $\rho_c^i = \frac{\lambda_c^i(\vec{n})}{\Gamma_c^i + \mu_c}$, $\vec{n} \in \mathbb{E}$ et G est la constante de normalisation donnée par

$$G = \sum_{\vec{n} \in \mathbb{E}} \prod_{i=1}^r \frac{(\rho_{RT}^i)^{n_{RT}^i}}{n_{RT}^i!} \frac{(\rho_{BE}^i)^{n_{BE}^i}}{n_{BE}^i!}.$$

Nous remarquons que cette probabilité dépend de la valeur moyenne du temps de séjour de l'utilisateur demandant le service de classe- c . Ce temps quant à lui, est en fonction de la probabilité d'avoir un appel de cette classe dans la région i et elle même dépend au temps moyen de séjour de l'utilisateur Γ_c^i dont lequel il dépend de la probabilité p_c^i dans l'équation (5.16). Alors encore cette dernière est en fonction de la distribution stationnaire et vice versa. Par conséquent, pour résoudre ce problème nous utilisons le théorème du point fixe. L'algorithme suivant nous permet de trouver la solution.

Algorithm 4 : Convergence des probabilités

- 1: Initialiser la probabilité dans l'équation (5.15) : $p_{c,old}^i = p_c^i = 0.1$.
 - 2: Calculer le temps moyen de séjour dans (5.15).
 - 3: Trouver la distribution de probabilités $\pi_1(\vec{n})$ via l'équation (5.21).
 - 4: Dédire la nouvelle valeur de la probabilité (5.16), notée par $p_{c,new}^i$.
 - 5: Vérifier la convergence de probabilité entre l'ancien et la nouvelle valeur, c'est à dire si $|p_{c,new}^i - p_{c,old}^i| < \zeta$, où ζ est un nombre très petit et strictement positive, alors la nouvelle probabilité est utilisée dans le calcul des performance du système. Sinon, passer à l'étape 2 avec la nouvelle valeur comme valeur initiale. Les itérations sont poursuivies jusqu'à atteindre la convergence des probabilités.
-

5.3.6 Analyse du système avec le deuxième mécanisme

Taux binaires par classe et espace d'états : Dans cette partie, nous analysons le système avec le mécanisme du contrôle d'admission décrit dans la figure 5.11. Nous considérons que l'opérateur réserve une capacité minimale pour les appels BE ainsi une portion de sous porteuses aux appels RT demandant la migration due à la mobilité. Notons respectivement ces capacités réservées par L_{BE} et L_{RT}^m . La capacité restante aux appels RT est notée par L_{RT} dont la forme est

$$L_{RT} = L - L_{BE} - L_{RT}^m. \quad (5.22)$$

Pour l'état \vec{n} du système, les ressources occupées par les appels temps réel à cet état sont notées par $B_2(\vec{n})$ et calculées par

$$B_2(\vec{n}) = \sum_{i=1}^r n_{RT}^i L_{RT}^i, \quad (5.23)$$

où

$$L_{RT}^i = \frac{R_{RT}}{k \times B \times e_i \times (1 - BLER_i)}.$$

La qualité de service des appels temps réel est de maintenir un débit constant partout dans la surface couverte par la station de base. Alors qu'un appel de type *best effort* va recevoir un débit instantané dépendant de la région i où se trouve le mobile noté R_{BE}^i . La capacité dynamique totale partagée équitablement selon le processus de partage PS entre tous les appels *best effort* notée $C(\vec{n})$, est définie par

$$C(\vec{n}) = \begin{cases} L - B_2(\vec{n}) - L_{RT}^m, & \text{si } B_2(\vec{n}) < L_{RT}; \\ L_{BE}, & \text{sinon.} \end{cases} \quad (5.24)$$

Donc le nombre de sous porteuses allouées à un appel de type *best effort* dans la région i est donné par la relation

$$L_{BE}^i(\vec{n}) = \left\lfloor \frac{C(\vec{n})}{\sum_{i=1}^r n_{BE}^i} \right\rfloor, \quad (5.25)$$

où $|x|$ représente la partie entière inférieure où égale à x . Le taux binaire atteint par un mobile dans la région i est $R_{BE}^i(\vec{n}) = L_{BE}^i(\vec{n}) \times k \times B \times e_i \times (1 - BLER_i)$. Par conséquent, la durée moyenne d'un appel *best effort* dans la région i est en fonction de l'état du système et la quantité moyenne à transmettre

$$\frac{1}{\mu_{BE}^i} = \frac{E(Pay)}{R_{BE}^i(\vec{n})}.$$

Le système accepte sans limite les appels de la classe *best effort* et donc nous définissons l'espace d'états en prenant en compte que les appels temps réel.

$$\mathbb{F} = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r n_{RT}^i L_{RT}^i \leq L_{RT} \}. \quad (5.26)$$

Par la suite nous avons besoin de la fonction indicatrice suivante

$$\delta(X) = \begin{cases} 1, & \text{si } X \text{ est vraie;} \\ 0, & \text{sinon.} \end{cases}$$

Taux de transition : Les taux de transition possibles du système à partir de l'état \vec{n} vers d'autres états sont obtenus comme suit. Soit \vec{n}_{i+}^c caractérise l'état du système quand un nouvel appel de classe- c est arrivé dans la région i en notant le taux transition correspondant par $q(\vec{n}, \vec{n}_{i+}^c)$. Soit \vec{n}_{i-}^c l'état déclenché quand un appel de classe- c dans la région i termine son service ou bien change sa modulation (région) et notons le taux de transition associé par $q(\vec{n}, \vec{n}_{i-}^c)$. Soit \vec{n}_{ij}^c ($j = i \pm 1$) un état correspondant au déplacement d'un appel de classe- c dans la région i vers la région voisine, et notons le taux de transition par $q(\vec{n}, \vec{n}_{ij}^c)$. En conséquent, nous formulons les taux de transition possibles comme suit

$$\begin{aligned} q(\vec{n}, \vec{n}_{i+}^{RT}) &= \delta(B_2(\vec{n}) + L_{RT}^i \leq L_{RT}) \lambda_{RT,i}^0, \\ q(\vec{n}, \vec{n}_{i+}^{BE}) &= \lambda_{BE,i}^0, \\ q(\vec{n}, \vec{n}_{ij}^{RT}) &= \delta(B_2(\vec{n}) + \Delta_{RT}^{ij} \leq L_{RT} + L_{RT}^m) \lambda_{RT}^{ij}, \\ q(\vec{n}, \vec{n}_{ij}^{BE}) &= \lambda_{BE}^{ij}, \\ q(\vec{n}, \vec{n}_{i-}^{RT}) &= n_{RT}^i (\mu_{RT} + \Gamma_{RT}^i), \\ q(\vec{n}, \vec{n}_{i-}^{BE}) &= n_{BE}^i (\mu_{BE}^i(\vec{n}) + \Gamma_{BE}^i), \end{aligned}$$

où $\Delta_{RT}^{ij} = L_{RT}^i - L_{RT}^j$ et Γ_c^i est calculé en utilisant l'Algorithme 4, en remplaçant la probabilité π_1 par π_2 . Une fois nous avons défini les transitions possible du système, nous construisons la matrice de transition notée \mathbf{Q} . Nous avons la relation $\mathbf{Q} = (q(\vec{n}, \vec{n}'))$ pour tout $\vec{n} \in \mathbb{F}$ et $\vec{n}' \in \mathbb{F}$. Le taux de transition de l'état \vec{n} à \vec{n}' est noté par $q(\vec{n}, \vec{n}')$. Ce taux devrait être calculé sachant que la somme de tous les termes de chaque ligne dans la matrice \mathbf{Q} et pour toutes les régions $i = 1, \dots, r$ et toutes les classes soit nulle.

Distribution stationnaire dans le cas 2 : Soient $\pi_2(\vec{n})$ la probabilité d'état d'équilibre du système à l'état \vec{n} ($\vec{n} \in \mathbb{F}$) et $\vec{\pi}$ le vecteur des probabilités d'équilibres de la forme $\vec{\pi} = \{\pi_2(\vec{n}) \mid \vec{n} \in \mathbb{F}\}$. Ce vecteur est donné en résolvant le système suivant

$$\vec{\pi} \mathbf{Q} = \vec{0}, \quad (5.27)$$

$$\vec{\pi} \vec{1} = 1. \quad (5.28)$$

où $\vec{1}$ est un vecteur colonne des unes et $\vec{0}$ un vecteur ligne des zéros.

5.3.7 Evaluation des performances

Après avoir trouver les distributions stationnaires correspondantes aux mécanismes de contrôle d'admission cités avant, nous fournissons dans cette partie des métriques de performance intéressantes : débit total du système, probabilités de blocage et de rejet, et temps de séjour.

Probabilités de blocage : La probabilité de blocage d'un appel dans la région i de la classe- c dans le cas du premier mécanisme, est donnée en considérant tous les états du système qui mènent vers un état bloquant dans la région i , par

$$B_c^i = \sum_{\vec{n} \in \mathbb{E}_c^i} \pi_1(\vec{n}), \quad i = 1, \dots, r, \quad (5.29)$$

où

$$\mathbb{E}_c^i = \{\vec{n} \in \mathbb{E} \mid B_1(\vec{n}) + L_c^i > L_0\}.$$

La probabilité de blocage d'un appel dans la région i de la classe temps réel dans le cas du deuxième mécanisme, est donnée par tous les états du système qui mènent vers un état bloquant dans la région i , par

$$B_{RT}^i = \sum_{\vec{n} \in \mathbb{F}_{RT}^i} \pi_2(\vec{n}), \quad i = 1, \dots, r, \quad (5.30)$$

où

$$\mathbb{F}_{RT}^i = \{\vec{n} \in \mathbb{F} \mid B_2(\vec{n}) + L_{RT}^i > L_{RT}\}.$$

Probabilités de perte : Un appel de classe- c en cours de communication dans la région i part vers la région j à cause de la mobilité du mobile dans le cas du premier mécanisme, est rejeté par le système par la probabilité suivante

$$D_c^i = \sum_{\vec{n} \in \mathbb{E}_c^{i,j}} \pi_1(\vec{n}), \quad i = 2, \dots, r, \quad (5.31)$$

où

$$\mathbb{E}_c^{i,j} = \{\vec{n} \in \mathbb{E} \mid B_1(\vec{n}) + L_c^j - L_c^i > L\}.$$

Un appel temps réel en cours de communication dans la région i part vers la région j à cause de la mobilité du mobile dans le cas du deuxième mécanisme, est rejeté par le système avec la probabilité

$$D_{RT}^i = \sum_{\vec{n} \in \mathbb{F}_{RT}^{i,j}} \pi_2(\vec{n}), \quad i = 2, \dots, r, \quad (5.32)$$

où

$$\mathbb{F}_{RT}^{i,j} = \{ \vec{n} \in \mathbb{F} \mid B_2(\vec{n}) + L_{RT}^j - L_{RT}^i > L_{RT} + L_{RT}^m \}.$$

Débit total moyen de la cellule : Dans le cas du premier mécanisme, le débit total moyen du système est calculé en balayant tous les états possibles du système

$$Th_1 = \sum_{\vec{n} \in \mathbb{E}} \pi_1(\vec{n}) \sum_{i=1}^r (n_{RT}^i R_{RT} + n_{BE}^i R_{BE}). \quad (5.33)$$

Ainsi que dans le cas du deuxième mécanisme, le débit total du système est

$$Th_2 = \sum_{\vec{n} \in \mathbb{F}} \pi_2(\vec{n}) \sum_{i=1}^r (n_{RT}^i R_{RT} + n_{BE}^i R_{BE}^i(\vec{n})). \quad (5.34)$$

Temps de séjour des appels *best effort* : Le temps moyen de séjour des appels *best effort* est

$$T^{BE} = \frac{\sum_{\vec{n} \in \mathbb{F}} \sum_{i=1}^r n_{BE}^i \pi_2(\vec{n})}{\sum_{i=1}^r \lambda_{BE,i}^0}. \quad (5.35)$$

5.3.8 Applications numériques

Dans cette partie, nous fournissons quelques résultats numériques concernant l'impact de nos mécanismes cités auparavant. Durant cette analyse nous considérons les paramètres suivants :

- La cellule OFDMA IEEE802.16e est décomposée en deux régions concentriques ($r = 2$) dont les rayons sont $R_1 = 300m$ et $R_2 = 600m$.
- Les modulations de ces régions sont 16-QAM 3/4 ($e_1=3$ bits/symbole) et QPSK 1/2 ($e_2=1$ bit/symbole) (Liu et al., 2005);
- La vitesse de modulation en *baud* est $B = 2666$ symbole/s;
- le taux d'erreur bloc $BLER_i = 0$;
- le nombre de fréquences est $K = 48$;
- le débit binaire d'appels RT R_{RT} est de 128 Kbps et celui d'appels BE R_{BE} est de 384 Kbps (Tarhini et Chahed, 2007b);
- la capacité totale du système L est de 10 porteuses;
- la durée moyenne d'appel RT est de 120s et la taille moyenne d'un fichier de téléchargement correspond à $E(Pay) = 5$ Mbits.

Enfin, nous supposons que les mobiles ou les utilisateurs se déplacent selon le processus RWP dans le disque convexe de rayon $R_z = 900m$. Ils choisissent aléatoirement des vitesses nouvelles dans chaque *waypoint* à partir d'une distribution uniforme de vitesse dans l'intervalle $[v_{min}, v_{max}]$, où $v_{min} = 3km/h$ (mobilité faible) ou bien $v_{min} = 20km/h$ (mobilité élevée), $v_{max} = 90km/h$.

Impact du premier mécanisme sur les performances du système : La figure 5.14 représente la probabilité de blocage pour chaque *burst profile* (modulation). Ces résultats sont exprimés en fonction de la partie des ressources réservées à la mobilité des utilisateurs. Comme prévu durant notre analyse théorique, les probabilités augmentent à mesure que le seuil réservé L_m augmente et l'efficacité de modulation diminue. D'ailleurs, une différence appréciable existe entre la probabilité de blocage des appels de la classe temps réel et ceux de la classe *best effort*. Ceci est dû à la bande passante demandée par type d'appel, un appel de classe RT dans notre environnement numérique exige plus de ressources qu'un appel BE. Ainsi, quand le seuil L_m augmente, les probabilités de blocages augmentent du fait de notre mécanisme de CAC, qui donne plus de priorité aux appels migrants que les nouveaux appels. Ainsi la probabilité de blocage dépend principalement de la bande passante et aussi de l'efficacité de modulation d'appel. Nous observons également avec l'augmentation du seuil L_m , les appels sont plus souvent bloqués quand ils viennent d'une région plus loin de la station de base (QPSK), et ils demandent plus de ressources dans ces régions. En particulier, nous observons sur cette figure les probabilités de blocage pour les deux types d'arrivées : BE-16QAM et RT-QPSK. Les probabilités de blocage sont exactement identiques parce que le produit entre la bande passante et l'efficacité de modulation est le même pour tous les deux appels.

Ensuite, la figure 5.15 montre le débit total de la cellule en fonction du seuil réservé à la mobilité. Sur cette figure deux comportements singuliers doivent être étudiés. La première observation principale traite l'augmentation du débit en raison d'une mobilité plus élevée. Quand le mobile se déplace plus rapidement, il augmente sa probabilité de changement de région (et alors son efficacité de modulation) par unité de temps. Ceci implique également que plus d'appels seront perdus en raison du manque de ressources. En effet, le système va implicitement forcer les appels à employer de meilleures modulations. Cette explication décrit comment une mobilité plus élevée permet d'atteindre un débit important. Notons que cette remarque est confirmée dans le réseau WiMAX (Peyre et El-Azouzi, 2009) ainsi que dans les réseaux ad-hoc (Grossglauser et al., 2002). D'autre part, nous pouvons critiquer dans la figure 5.15, l'impact de la réservation de ressources pour faciliter la mobilité des appels. Pour n'importe quel comportement ou modèle de mobilité, le débit total moyen de la cellule diminue à mesure que la partie réservée augmente. Mais la chute du débit dépend du comportement de mobilité. Certainement, l'introduction des ressources réservées, nous aide de plus en plus à avoir des migrations d'appels dans les régions éloignées de la station de base. Ces appels utilisent plus de sous-canaux pour atteindre le même débit binaire que celui recevant un mobile de la première région. En conséquence, le système atteint un faible débit en donnant la priorité à la gestion de mobilité d'utilisateur.

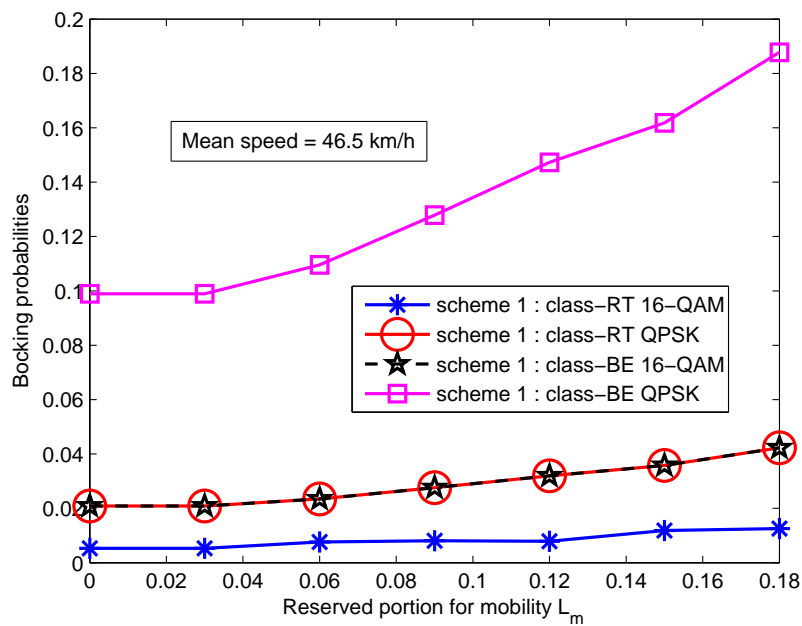


FIGURE 5.14 – Probabilités de blocages en fonction du seuil L_m et la vitesse moyenne pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

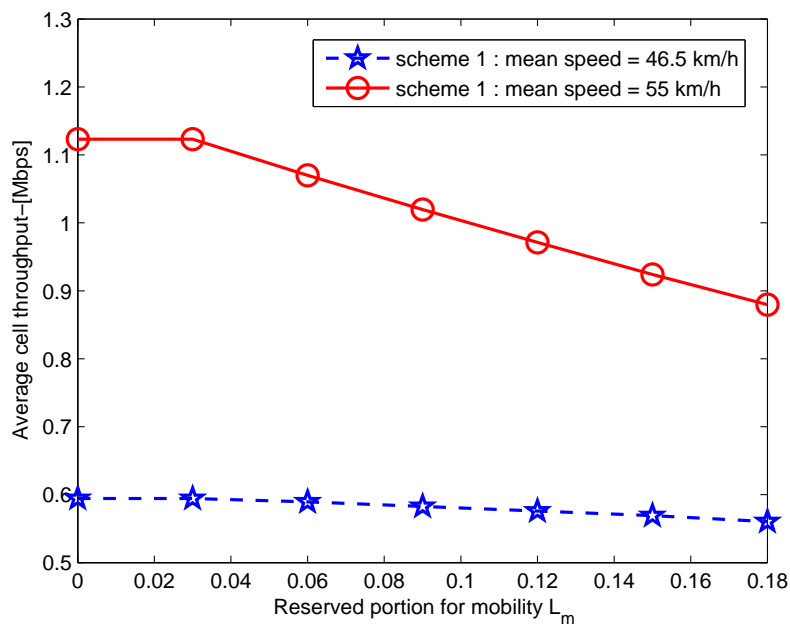


FIGURE 5.15 – Débit total moyen en fonction du seuil L_m et la vitesse moyenne pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

Par contre plus l'impact de la première stratégie, nous représentons dans la figure 5.16 les probabilités de pertes d'appels en fonction du seuil L_m . Nous traçons les résultats pour les deux types de trafics migrant de la première région (c'est à dire modulation 16-QAM) vers la région frontalière (c'est à dire de modulation QPSK), et pour deux comportements de mobilité. La figure montre le fort impact de la réservation des ressources sur la gestion de mobilité. Pour ce qui concerne les trafics temps réel, nous observons qu'une mobilité plus élevée cause une augmentation 20 fois supérieures des pertes d'appels. Pour lutter contre cet effet, notre mécanisme de CAC permet la réservation de ressources aux appels en mobilité afin d'avoir une probabilité de perte d'appel inférieure à une valeur désirée. Par exemple, pour une réservation de ressources inférieure à dix pour cent de la capacité totale du système, nous observons une diminution de la probabilité de perte à moins d'un-pour cent. Pour les appels *best effort*, nous constatons exactement le même comportement. Néanmoins, la probabilité de perte des trafics BE demeure plus élevée que celle des appels RT, par ce que les appels BE demandent un débit binaire plus grand que celui des appels RT. Donc, en demandant plus de ressources pour atteindre leur débit, ils subissent plus de perte. Pour ce type de trafic, notre stratégie de CAC permet également de réduire considérablement la probabilité de perte en augmentant la réservation de ressources L_m .

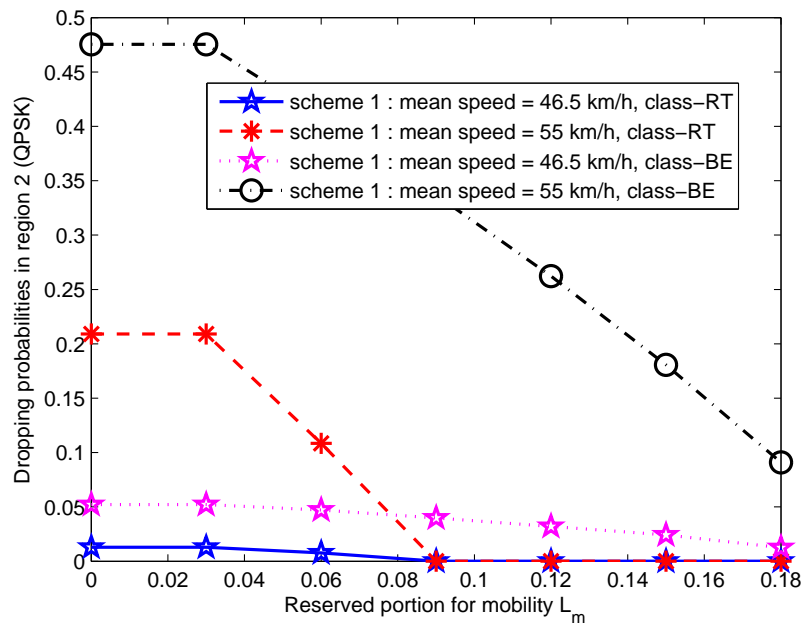


FIGURE 5.16 – Probabilités de perte en fonction du seuil L_m et la vitesse moyenne pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

Impact du deuxième mécanisme sur les performances du système : La figure 5.17 montre les probabilités de blocage pour les trafics temps réel dans les deux régions en fonction de la réservation de ressources pour les appels *best effort*. Nous traçons les

résultats obtenus pour deux profils de réservation de ressource pour la gestion de mobilité d'utilisateur intra cellule. Cette figure montre clairement l'impact de L_{BE} et L_{RT}^m sur les probabilités de blocage dans chaque région. Généralement, les probabilités de blocage sont fortement augmentées quand la réservation minimale de ressources aux appels BE augmente. En outre, cet effet devient plus important par l'augmentation de la réservation de ressource de gestion de mobilité. De ce rapport, nous pourrions calculer les gammes de valeurs pour L_{BE} et L_{RT}^m qui satisfont un seuil de blocage maximum de probabilité. La figure 5.18 montre les probabilités de perte pour les trafics temps réel

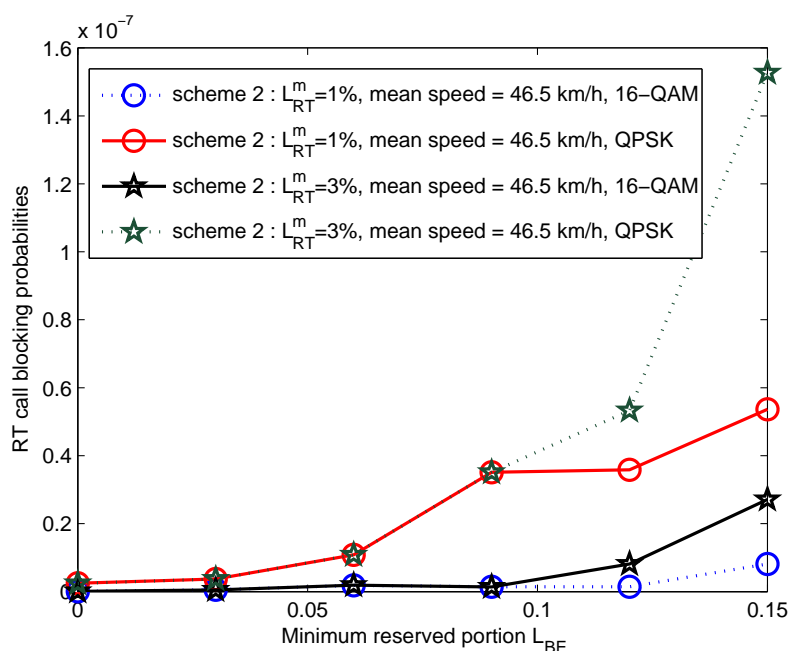


FIGURE 5.17 – Probabilités de blocages des appels RT en fonction du seuil L_{BE} et L_{RT}^m pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

dans la région frontalière en fonction de la réservation minimale de ressources pour les appels *best effort*. Nous traçons les résultats obtenus pour deux profils de réservation de ressource pour la gestion de la mobilité utilisateur. La figure montre comment la réservation de ressources L_{RT}^m diminue la probabilité de perte via le seuil minimal réservé aux appels BE. Dans cette figure, la probabilité de perte est réduite quand la réservation L_{BE} est supérieure à dix pour cent. Ainsi nous pouvons déterminer les valeurs possibles pour L_{RT}^m arrivant à une probabilité de perte d'appels RT maximum désirée.

Enfin, la dernière figure 5.19 présente le débit total de la cellule. Cette figure représente l'impact des différentes réservations de ressources pour la mobilité sur l'augmentation d'appels BE et la facilité de la mobilité des appels. Dans la figure 5.19 nous pouvons apprécier l'important impact de la réservation L_{BE} . En effet, par la réservation de 15% de la capacité totale pour les appels BE, nous doublons le débit total moyen de la cellule. En plus, l'augmentation de la réservation L_{RT}^m de mobilité diminue légèrement le débit total. Cette observation mène à constater qu'une petite quantité de sous-canaux

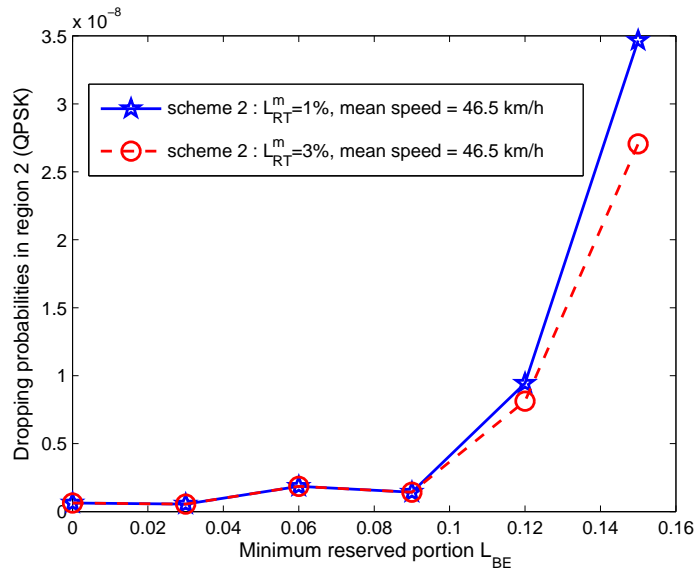


FIGURE 5.18 – Probabilités de perte d’appels RT en fonction du seuil L_{BE} et L_{RT}^m pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

réservés pour les appels BE permet d’atteindre un débit très élevé. En fait, en réservant quelques sous-canaux aux appels BE, nous augmentons également les probabilités de blocage et de perte d’appels RT. Par conséquent, ces quantités de ressources libérées par les appels bloqués ou perdus permettent aux appels BE mieux utiliser les ressources.

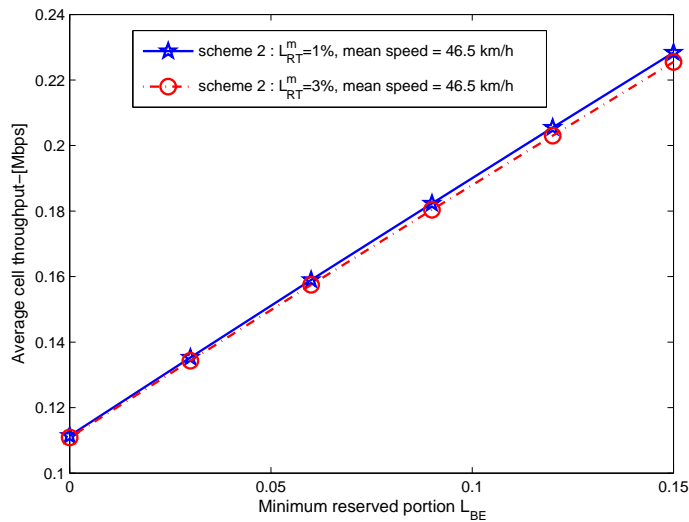


FIGURE 5.19 – Débit total moyen en fonction du seuil L_{BE} et L_{RT}^m pour $\lambda_{RT,i}^0 = \lambda_{BE,i}^0 = 0.3$.

5.4 Conclusion

Dans le cas discret du système IEEE802.16e, nous avons défini un modèle pour une cellule IEEE802.16e. Ce modèle se base sur la décomposition de la cellule en différentes régions de modulation suivant le principe de la technique du codage et modulation adaptative AMC. Nous avons proposé un nouveau mécanisme de contrôle d'admission qui fait une différenciation des les deux principaux types de service, temps réel et non temps réel. Les appels non temps réel ont des ressources minimums à partager équitablement *process sharing* entre eux, ainsi que toute capacité laissée par les appels de la première classe. Ainsi les appels temps réel ont plus de priorité par ce mécanisme en fonction de la gestion de ressources total du système. Ces derniers reçoivent un débit fixe tout le temps durant lequel les mobiles se déplacent dans toute la surface de la cellule. Alors que les appels non temps réel auront un débit qui varie selon la région où se trouve le mobile et la capacité restante par les appels temps réel. Ce mécanisme de contrôle d'admission facilite la gestion des ressources en rejetant un nouvel appel temps réel qui demande des ressources où le système à cet instant détecte qu'il existe une région bloquante intérieure à sa région, sans vérifier s'il y a des ressources libres. En plus de ce mécanisme, nous avons étudié l'impact du principe de partitionnement des codes entre les appels temps réel et non temps réel. Les appels temps réel disposent d'une plus grande plage de codes pour envoyer leurs requêtes de ressources. Aussi, une plage de code est partagée entre les appels de ces deux grandes classes de services.

Dans le cas continu du système IEEE802.16e, nous avons proposé deux mécanismes de contrôle d'admission ou stratégies de gestion de QoS dans le système IEEE802.16e. Le premier définit des débits binaires constants (CBR) pour les deux classes de service, temps réel et *best effort*. Nous avons présenté aussi une réservation de ressource pour faciliter la mobilité des utilisateurs entre les régions de la cellule. Le deuxième mécanisme garde la même qualité de service pour les appels temps réel et remplace celle des *best effort* par un partage équitable de ressources via le processus (PS), et nous avons réservé une bande passante minimale pour les appels *best effort* afin de leurs garantir une QoS minimale et une autre pour les appels temps réel en mobilité intra cellule. D'ailleurs, nous avons défini un modèle plus réaliste de mobilité RWP qui nous a permis de calculer les taux migration des appels de manière explicite. En nous basant sur ces propositions nous modélisons notre système dans les deux stratégies cités ci-dessus avec une chaîne de Markov à temps continue pour trouver la distribution stationnaire du système. Par les résultats numériques nous avons pu voir l'impact de ces deux stratégies sur les performances du système. Le choix de la stratégie dépend directement de la façon dont l'opérateur veut gérer son réseau par rapport à la qualité de service qu'il peut offrir aux ses clients. Comme perspective de ce travail, nous pensons utiliser le modèle de mobilité RWP également dans le cas discret et ainsi considérer le temps de pause (*thinking time*) d'un mobile dans sa région afin de décider du changement de son efficacité de modulation ou bien jusqu'à une autre région.

Conclusion générale et perspectives

Nous présentons à ce niveau une synthèse globale du manuscrit. Dans le but de situer nos travaux dans le contexte général, le premier chapitre présente un rappel de quelques concepts généraux portant sur les réseaux de nouvelles générations étudiés dans cette thèse tels que WCDMA, HSDPA et IEEE802.16e. Egalement, il consiste en un résumé des travaux antérieurs.

Nous nous sommes intéressés dans la première partie à l'amélioration des performances du lien montant du système WCDMA. Dans le chapitre 2 de cette partie, nous avons effectué une analyse de l'évaluation de la capacité en nous basant sur l'approche de l'analyse spectrale en présence d'une part de deux classes de services (temps réel et *best effort*) et d'autre part de trois classes de services (temps réel, non temps réel et *best effort*). Le cas de deux classes nous permet d'améliorer la gestion de la qualité de service en réduisant la complexité d'analyse des différentes classes de trafic via l'analyse spectrale. Ce qui nous a permis de trouver explicitement la distribution d'états d'équilibre du système dans deux scénarios de nombre d'appels *best effort* (fini et infini). De plus, nous avons validé nos résultats par le biais du simulateur NS2. Étant donné que certains services de classe *best effort* ont besoin d'un minimum de taux de transmission à garantir par l'opérateur, nous avons défini une nouvelle classe de services intermédiaire entre les deux classes étudiées dans la littérature temps réel et *best effort*. Cette classe est appelée non temps réel. Elle regroupe tous les services qui demandent un débit minimum comme le protocole de transfert de fichiers (FTP). En outre, le système WCDMA est modélisé à nouveau, en considérant les trois classes. En effet l'avantage de la méthode de l'analyse spectrale est de réduire la complexité du calcul. Les performances du système ne dépendent pas du nombre d'appels *best effort*. Cette indépendance nous permet de ramener le problème avec un nombre arbitraire d'appels *best effort* (voir infini) à un problème relativement simple qui ne dépend que des paramètres du système. Cependant, quand le nombre d'appels temps réel augmente, la méthode prend un temps de calcul important et peut donner quelques erreurs, ce qui nous motive à utiliser une autre approche complémentaire et rapide au niveau temps de calcul. Dans le chapitre 3, nous présentons cette approche basée sur la perturbation singulière afin d'approcher la distribution d'états d'équilibre du système WCDMA. Nous développons la distribution en série de Taylor dépendant du paramètre de perturbation ϵ . Nous montrons que le premier coefficient de la série de Taylor donne déjà une bonne approximation de la distribution stationnaire avec moins de calcul comparé à l'analyse spectrale. Le fournisseur de service peut utiliser cette approche pour évaluer en **temps**

réel les performances du système.

Dans la deuxième partie de ce manuscrit, nous nous sommes intéressées à la gestion des ressources radio des réseaux HSDPA et IEEE802.16e en lien descendant en tenant compte de la mobilité des utilisateurs. Dans le chapitre 4, nous nous sommes focalisé sur l'analyse du système HSDPA comme extension du réseau UMTS, et avons présenté une approche pour servir les utilisateurs mobiles avec un débit constant même s'ils changent leur efficacité de modulation, par exemple, en passant de la haute modulation 64-QAM à la faible QPSK ou inversement. Dans le réseau haut débit HSDPA, l'efficacité de modulation change généralement en fonction de la qualité du canal. En d'autres termes, une bonne qualité de signal permet d'affecter à l'utilisateur une haute modulation avec plusieurs bits par symbole et inversement. Cependant, quand la modulation change de la haute valeur vers la faible, l'utilisateur se trouve dans un débit dégradé. Ce problème est traité dans le chapitre 4 en ayant recours à l'affectation dynamique de la bande passante quand la modulation change et ce dans le but de maintenir un débit constant à l'utilisateur indépendamment de sa position. La cellule est subdivisée en un nombre fini de régions caractérisées chacune par une modulation donnée. Un nouveau contrôle d'admission est proposé opérant avec deux classes de services temps réel et non temps réel avec différents scénarios de mobilité des utilisateurs. Le contrôle d'admission gère les ressources du système en prenant en compte la mobilité des utilisateurs aussi bien entre les régions qu'entre les cellules. Ce contrôle d'admission ne donne aucune priorité aux classes d'appels considérées et par contre, favorise les appels en mobilité sur les nouveaux appels en leur réservant certaines ressources. La réservation de ressources dépend de ce que veut faire un opérateur pour la gestion du trafic de son réseau en fonction de ses besoins et ses exigences. L'opérateur peut changer la valeur réservée à la gestion de mobilité des utilisateurs selon les périodes où la mobilité est forte ou faible. Nous avons montré que dans le cas de mobilité inter cellules, l'interférence intercellulaire oblige notre mécanisme du CAC à allouer plus de bande passante aux mobiles pour obtenir un débit constant. Ce qui réduit l'espace d'états possibles du système et donc le débit total du système se voit dégrader. Ces résultats permettent de trouver un bon compromis entre les performances globales du système en termes de débit total, de probabilités de perte et de blocage.

Ensuite, dans le chapitre 5, nous nous sommes intéressés à la qualité de service offert par le système IEEE802.16e. Nous avons étudié, dans un premier volet, le cas discret du système IEEE802.16e. Nous avons défini un modèle pour une cellule IEEE802.16e en nous basant sur la décomposition de la cellule en différentes régions de modulation suivant le principe de la technique du codage et modulation adaptatifs AMC. Nous avons modélisé le système par une chaîne de Markov à temps discret CMTD et avons proposé un nouveau mécanisme de contrôle d'admission qui différencie les deux principales classes de service, temps réel (RT) et non temps réel (NRT). Les appels non temps réel ont des ressources minimales à partager équitablement entre eux, ainsi que toute capacité laissée par les appels de la première classe. En effet, les appels temps réel ont plus de priorité tenue en compte par ce mécanisme. Ces derniers reçoivent un débit constant même lorsque les mobiles se déplacent dans la surface de la cellule. Quant aux appels non temps réel, ils auront un débit qui varie selon la région et la capacité

restante. Le mécanisme de contrôle d'admission développé dans ce travail facilite la gestion des ressources en rejetant un nouvel appel temps réel qui demande des ressources où le système à cet instant détecte qu'il existe une région bloquante intérieure à sa région, sans vérifier s'il y a des ressources libres. En outre, nous avons étudié l'impact du principe de partitionnement des codes entre les appels temps réel et non temps réel. Les appels temps réel disposent d'une plus grande plage de codes pour envoyer leurs requêtes de ressources. Aussi, une plage de code est partagée entre les appels de ces deux grandes classes de services. Le couplage du mécanisme du CAC et de la politique de partitionnement des codes de connexions entre les utilisateurs permet au fournisseur de service de mieux gérer et d'optimiser la gestion de la capacité de son réseau. En effet, s'il n'y a pas de ressources libres parmi celles réservées aux appels temps réel, la station de base n'assigne aucun code à un nouvel appel temps réel, et donc l'allouera à un appel de type non temps réel. En proposant un partitionnement dynamique des codes de connexions, il devient possible d'adapter l'espace des codes d'une classe de service en fonction des ressources disponibles. Ainsi, une classe de service pour laquelle peu de ressources sont disponibles verrait sa plage de codes réduite. De cette manière, peu de requêtes de ce type aboutiront au niveau de la station de base. En conséquence, la longueur de la file d'attente des requêtes sera réduite, diminuant ainsi le délai de traitement de l'ensemble des requêtes.

Dans le deuxième volet du chapitre 5, nous avons étudié le cas continu du système IEEE802.16e, nous avons proposé deux algorithmes de contrôle d'admission ou stratégies de gestion de QoS dans le système IEEE802.16e. Le premier définit des débits binaires constants pour les deux classes de service, temps réel et *best effort*. Une réservation de ressource pour faciliter la mobilité des utilisateurs entre les régions de la cellule est aussi proposée. Le deuxième mécanisme garde la même qualité de service pour les appels temps réel et remplace celle des *best effort* par un partage équitable de ressources, tout en réservant une bande passante minimale pour les appels *best effort* afin de leur garantir une QoS minimale et une autre pour les appels temps réel en mobilité intra cellule. D'ailleurs, nous avons implémenté un modèle plus réaliste de mobilité qui s'appelle RWP et qu'est souvent utilisé dans les réseaux ad-hoc. Il est introduit récemment dans les réseaux cellulaires par Hyytiä et al. dans (Hyytiä et Virtamo, 2007). Nous étendons le modèle RWP dans une cellule IEEE802.16e subdivisée en un nombre fini de régions. Ce modèle nous a permis de calculer les taux de migration d'appels entre les régions de la cellule de manière explicite. En nous basant sur ces propositions nous modélisons notre système à l'aide d'une chaîne de Markov à temps continue CMTC pour trouver la distribution stationnaire du système. A travers les résultats numériques nous avons pu voir l'impact de ces deux stratégies sur les performances du système. Le choix de la stratégie dépend directement du choix dont l'opérateur veut gérer son réseau par rapport à la qualité de service qu'il peut offrir aux ses clients.

A l'issue des travaux menés tout au long de cette thèse, un certain nombre de pistes de recherche méritent, à notre sens, d'être explorées. Nous les résumons dans les points suivants :

- dans les trois premières contributions nous avons considéré uniquement le lien montant du système WCDMA, il serait très intéressant de considérer le lien des-

endant du système avec l'impact des interférences inter cellules. Ainsi d'évaluer ces performances par les deux méthodes : celle d'analyse spectrale et celle de la perturbation singulière ;

- dans les deux dernières contributions nous avons axé notre recherche sur le lien descendant du réseau haut débit IEEE802.16e, il serait important de comparer les mécanismes de la stratégie de gestion de ressource pour déterminer celui qui est optimal pour un opérateur ;
- la considération de trois classes de trafics dans les réseaux HSPA et WiMAX pour définir un nouveau contrôle d'admission en garantissant une qualité de service uniquement pour les appels des deux premières classes ;
- la proposition d'un nouveau mécanisme de contrôle d'admission dans les réseaux HSPA et WiMAX (IEEE802.16e) joignant (ou combinant) les deux liens montant et descendant.

Liste des publications

OUVRAGES SCIENTIFIQUES

1- Title : CAC Algorithms based on Random Waypoint Mobility for IEEE802.16e Networks

Authors : [Khalil IBRAHIMI](#), Rachid EL-AZOUZI, Thierry PEYRE et El-Houssine BOUYAKHF.

Chapitre de 24 pages accepté pour la publications dans le livre " *Radio Communications* ", <http://intechweb.org/>, ISBN 978-953-7619-X-X, 2009.

REVUES INTERNATIONALES AVEC COMITE DE LECTURE

1- Title : WCDMA Uplink System Capacity : A Singular Perturbation Approach

Authors : [Khalil IBRAHIMI](#), Rachid EL-AZOUZI, Abdellatif KOBANE et El-Houssine BOUYAKHF.

16 pages publiées dans la revue *Applied Mathematical Sciences (AMS)*, Vol. 2, 2008, no. 32, 1581-1596, Avril 2008.

Abstract : *We consider a WCDMA system with two types of calls : real time (RT) calls that have dedicated resources, and data non real time (BE) calls (i.e. best effort) that share system capacity. We consider reservation of some capacity resources for the BE traffic as well as any capacity left over from RT calls. Our analysis approach is based on modeling of the system as a two dimensional Markov chain, where the first correspond to the number of RT calls and the second to the number of BE calls in the system. In order obtain the steady state distribution of this system, we use a singular perturbation solution approach for approximating the steady state. Our approach gives a good approximation and relatively faster computation in comparison with our exact methods like spectral analysis. Based on this analysis, we derive performance evaluation results regarding blocking of RT calls and sojourn time of BE calls, under different traffic characteristics. Finally, we extend our result to cover BE admission control. Here, we find the maximum BE arrival rate such that the sojourn time of BE calls not exceed a threshold.*

2- Title : The Uplink Capacity Evaluation of Wireless Networks : Spectral Analysis Approach

Authors : Abdellatif KOBANE, Rachid EL-AZOUZI, [Khalil IBRAHIMI](#), Sujit Samanta et El-Houssine BOUYAKHF.

14 pages acceptées pour la publications dans la revue *Computing and Information Technology (CIT)*, 2009.

Abstract : *In this paper we study the capacity of wireless cellular network, in particular, we focus our study on the WCDMA system. Considering two types of calls : real-time (RT) calls characterized by a quasi fixed transmission rate and best-effort (BE) calls which do not require strict demand (but need some reliability conditions). In this paper we develop an approach based on spectral analysis for evaluating the cell capacity. We explicitly obtain the simultaneous distribution of the number of RT connections and the number of BE connections in the steady-state. This analysis allows us to simplify the computation of the performance measures including expected delay and throughput of BE traffic. These performances are obtained explicitly in both cases (finite and infinite) of BE calls as function of system parameters like arrival rate of BE and RT calls, service rate of BE and RT calls. These results allow the operator to evaluate the cell capacity by varying these parameters independently of the number of BE calls according to its policy to manage the network. Note that this analysis can be applied to various systems as WiMAX/HSPA and for both uplink and downlink scenarios, so our spectral analysis approach is not exclusively applicable to the uplink of WCDMA system. We further propose some CAC (Call admission control) policies for BE traffic. We finally conclude this work by some numerical and simulation results. The simulation results obtained by the network simulator (NS2) are closely to the numerical results which validate our theoretical model.*

3- Title : IEEE802.16e Cell Capacity including mobility management and QoS Differentiation

Authors : Thierry PEYRE, Rachid EL-AZOUZI et [Khalil IBRAHIMI](#).

10 pages soumis pour la publications dans la revue *Mobile Networks and Applications*, 2009.

Abstract : *In this paper we study the capacity of the OFDMA-based IEEE802.16 WiMAX network in the presence of two types of traffic : Real-Time and Non-Real-Time, including Adaptive Modulation and Coding (AMC) with user intra cell mobility. Our work deals with the performance of the connexion admission and ressource allocation algorithm over a general ranging arrival process : it has been recently proved that the exponential distribution is inappropriate. Based on the generalized traffic processes, we develop a resource allocation that maintains the bit rate of real time connections independently of the user position in the cell. Moreover, we enhance the CAC algorithm with a mecanism designed for the mobility management. Using a novel Discrete Time Markovian analysis we evaluate the impact of our resource allocation for mobility behavior on the RT and NRT connections as blocking and dropping probability, mean sojourn time and throughput.*

COMMUNICATIONS/CONFERENCES INTERNATIONALES AVEC ACTE

1- Title : Adaptive Modulation and Coding scheme with intra- and inter-cell mobility for HSDPA system

Authors : [Khalil IBRAHIMI](#), Rachid EL-AZOUZI, Sujit SAMANTA et El-Houssine BOUYA-KHF.

8 pages publiées dans la conférence *BROADNETS International Conference on Broadband Communications, Networks, and systems*, Septembre 2009.

Abstract : *The Adaptive Modulation and Coding (AMC) scheme which handles user's mobility issue plays a significant role to improve the desired quality of service in High Speed Downlink Packet Access (HSDPA) networks. We develop a resource allocation which maintains constant bit rate for real-time (RT) and non-real-time (NRT) calls independently of the user position in the cell. We derive analytical results relating to user performance in terms of blocking and dropping probabilities, and the results relating to the system performance in term of total throughput. We propose a call admission control scheme to derive a good trad-off between the blocking and dropping probabilities, and the choice of scheduling that efficiently utilizes the available resources.*

2- Title : CAC Algorithms based on Random Waypoint Mobility for IEEE802.16e Networks

Authors : [Khalil IBRAHIMI](#), Rachid EL-AZOUZI, Thierry PEYRE et El-Houssine BOUYAKHF.

8 pages publiées dans la conférence *IEEE NGNS'09 International Conference on Next Generation Networks and Services*, Juin 2009.

Abstract : *In this paper we propose the performances reached by two Connection Admission Control (CAC) algorithms for the OFDMA-based IEEE802.16e (WiMAX) network in presence of two types of traffic : Real-Time (RT) and Best-Effort (BE), including Adaptive Modulation and Coding (AMC). We consider intra mobility by using an realistic and general mobility model through the Random Waypoint (RWP) approach. In the first CAC scheme, both RT and BE calls receive the same bit rate independently of the mobile position in the cell. Whereas in the second proposal, only the RT calls receive the same bit rate whatever are the modulation efficiency (i.e. AMC region) whereas the BE calls fairly share the available resources using a Processor Sharing technic. In the same time, we introduce a reserving amount of resources in order to provide an high priority the migrating calls. Furthermore, we develop the system performances based on a continuous time approach and provide extensive numerical examples in term of : throughput, sojourn time, blocking and dropping probabilities.*

3- Title : Uplink Call Admission Control in Multi-services W-CDMA Network

Authors : [Khalil IBRAHIMI](#), Rachid EL-AZOUZI et El-Houssine BOUYAKHF.

6 pages publiées dans la conférence internationale *IEEE Symposium Computer and Communication (ISCC)*, Juillet 2008.

Abstract : *The capacity of CDMA wireless network is usually studied considering two classes of services : real-time and best-effort. In this paper, we are interested in analyzing sharing between three classes of services : real-time (RT), non-real-time (NRT) and best-effort (BE). A classical approach which is widely used in wireless networks is based on adaptively deciding how many channels to allocate to calls of a given class. The rational behind our idea is that the NRT class (e.g. FTP) requires a minimum transmission rate. The capacity allocated to NRT traffic includes a fixed portion of bandwidth as well as a dynamic part which is shared with RT service. In contrast, BE applications can adapt their transmission rate to the network's available resources.*

Hence, the best-effort applications can use only the unused resources of the NRT band. Using a spectral analysis approach, we compute the steady-state distribution of the calls number for those different classes which allows us to provide explicitly the performance measure. The QoS parameters of interest are primarily the blocking probability for both RT calls and NRT calls, and expected sojourn times for both NRT calls and BE calls. We finally provide numerical study to show the benefit of our capacity allocation method by providing a desired quality level of service for NRT services, and we propose some CAC policies for NRT and BE services.

4- Title : IEEE802.16 Multi-Class Capacity Including AMC Scheme and QoS Differentiation for Initial and Bandwidth Request Ranging

Authors : Thierry PEYRE, [Khalil IBRAHIMI](#) et Rachid ELAZOUZI.

9 pages publiées dans la conférence internationale *Performance Evaluation Methodologies and Tools (ACM/ValueTools)*, Octobre 2008.

Abstract : *In this paper we study the capacity of the OFDMA-based IEEE802.16 WiMAX network in the presence of two types of traffic, streaming (Real-Time) and elastic (Non-Real-Time) including Adaptive Modulation and Coding (AMC). Many studies in the literature assumed that packets or calls arrive to the system according to poisson process for the sake of analytical simplicity. However, it has been recently proved that the exponential distribution is inappropriate. Based on the generalized traffic processes developed, we study the media access control (MAC) layer of WiMAX and develop a resource allocation that maintain the bit rate of real time connections independently of the user position in the cell. Using Markovian analysis we evaluate the impact of our resource allocation on the non-real time connection (NRT) as expected delay and throughput.*

5- Title : Capacity Evaluation of Multi-Service W-CDMA : A Spectral Analysis Approach

Authors : Abdellatif KOBANE, Rachid EL-AZOUZI, [Khalil IBRAHIMI](#) et El-Houssine BOUYAKHF.

7 pages publiées dans *IASTED Communication Systems and Networks (CSN)*, Aout 2007.

Abstract : *We study the capacity of a multi-service WCDMA technology. We consider two types of calls : real time calls (RT) and non-real time calls as Best effort (BE). The first type of calls is characterized by a quasi static transmission rate. In this paper we develop an approach based on spectral analysis for assessing the cell capacity. We obtain explicitly the simultaneous distribution of number of RT connections and the number of BE connections in the steady-state. This analysis allow us to simplify the computation of blocking probability of the RT as well as the performance measures including expected delay and throughput. By calculating the blocking probability of BE cell, we further propose some CAC policies for BE traffic.*

COMMUNICATIONS NATIONALES

[K. Ibrahim](#), M. Baslam et S. Elhammani, Contrôle d'Admission du Trafic Non Temps Réel dans le réseau multiservices IEEE802.16e. Dans les actes de la MajecSTIC' 09, 16-18 Novembre 2009, Avignon, France.

H. Tembine, A. Kobbane, [K. Ibrahimi](#), Jeu de minorité appliqué au routage dans les réseaux, Dans les actes de JOSTIC 07 JOurnées Scientifiques en Technologie de l'Information et de la Communication, 24 février 2007, Rabat, Maroc.

[K. Ibrahimi](#) and E.H. Bouyakhf, Resource Allocation in OFDMA/WiMAX System, Dans les actes JOSTIC 08 JOurnées Scientifiques en Technologie de l'Information et de la Communication, 3-4 Novembre 2008, Rabat, Maroc.

Rapports internes de recherche(Projet MAROC TELECOM)

E.H. Bouyakhf, R. El-Azouzi, [K. Ibrahimi](#), E. Sabir, Ralph El-Khoury, "Capacité des réseaux UMTS et HSDPA avec des applications multimédias" ; Octobre 2008.

[K. Ibrahimi](#), R. Elazouzi, A. Kobbane, and E.H. Bouyakhf, "Capacity of CDMA cellular Network in Multiservices Environments" ; Octobre 2007.

A. Kobbane, R. El-Azouzi, [K. Ibrahimi](#), and E.H. Bouyakhf, "Evaluation of WCDMA Capacity Network", Avril 2007.

Liste des illustrations

1.1	Méthodes d'accès multiples	21
1.2	Etalement à séquence directe DS-CDMA	22
1.3	Trame OFDMA	23
1.4	Points de quanstellation	24
1.5	Evolution des réseaux sans fil	27
1.6	Architecure globale du réseau UMTS	29
1.7	Elements de réseau d'accès UMTS	30
1.8	Architecure système avec le HSDPA	31
2.1	Représentation de P_B^{RT} en fonction de L_{BE}	62
2.2	Temps de séjour d'appels BE en fonction du seuil	62
2.3	Temps de séjour vs seuil	63
2.4	Taux d'arrivée maximal en fonction du seuil du temps de séjour	63
2.5	Résultats analytiques et de simulation sur NS2 pour RT	64
2.6	Résultats analytiques et de simulation sur NS2 pour BE	64
2.7	Blocage d'appels RT	75
2.8	Tems de séjour d'appels NRT	76
2.9	Blocage d'appels NRT RT en fonction du seuil NRT	76
2.10	Temps de séjour BE RT en fonction du seuil NRT	76
2.11	Taux maximum NRT RT en fonction du seuil du temps de séjour	77
2.12	Taux maximum BE RT en fonction du seuil du temps de séjour	77
3.1	Probabilité en fonction du nombre d'appels BE	91
3.2	Temps de séjour en fonction du seuil L_{BE} , pour $\mu_{RT} = 10$	91
3.3	Temps de séjour en fonction du seuil L_{BE} , pour $\mu_{RT} = 1$	92
3.4	Temps de séjour en fonction du seuil L_{BE} , pour $\mu_{RT} = 0.1$	92
4.1	La cellule HSDPA sectorisée en plusieurs régions avec AMC	100
4.2	Probabilités de blocage en fonction du taux d'arrivée d'appels RT	110
4.3	Probabilités de blocage d'appels NRT en fonction du taux d'arrivée	111
4.4	Débit total en fonction du taux d'arrivée RT	112
4.5	Probabilités de blocage en fonction de la bande passante réservée	113
4.6	Probabilités de blocage en fonction de la bande passante réservée	114
4.7	Probabilités de perte en fonction de la bande passante réservée	114
4.8	Débit total en fonction de la bande passante réservée	115

4.9	Probabilité de perte en fonction de la bande passante réservée	115
4.10	Utilisation total de la bande passante en fonction du taux d'arrivée RT .	116
4.11	Probabilité de perte en fonction du taux d'arrivée RT	116
5.1	La cellule IEEE802.16e sectorisée en plusieurs régions avec AMC	122
5.2	Processus de connexion de <i>backoff</i> IEEE802.16e	122
5.3	La distribution des arrivées des connexions RT et NRT	123
5.4	Débit moyen des appels NRT en fonction de la capacité réservée	133
5.5	Probabilité de blocage des appels RT en fonction de la capacité réservée	133
5.6	Temps de séjour des appels NRT en fonction de la capacité réservée . . .	134
5.7	Débit moyen des appels NRT en fonction de la capacité réservée	134
5.8	Probabilité de blocage des appels RT en fonction de la capacité réservée	135
5.9	Temps de séjour des appels NRT en fonction de la capacité réservée . . .	135
5.10	Mécanisme 1	138
5.11	Mécanisme 2	138
5.12	La cellule OFDMA IEEE802.16e séctorisée.	140
5.13	Le domaine convexe du processus RWP	142
5.14	Probabilités de blocages en fonction du seuil L_m	150
5.15	Débit total en fonction du seuil L_m	150
5.16	Probabilités de perte en fonction du seuil L_m	151
5.17	Probabilités de blocages des appels RT en fonction du seuil L_{BE}	152
5.18	Probabilités de perte des appels RT en fonction du seuil L_{BE}	153
5.19	Débit total en fonction du seuil L_{BE}	153

Liste des tableaux

1.1	Modulations et taux de codages adaptatifs sur le IEEE802.16e	25
1.2	Historique du standard IEEE802.16	32
2.1	Valeurs numériques WCDMA	61
4.1	Partitionnement du SINR avec AMC	100
5.1	AMC selon IEEE802.16e	140

Liste des indexes

- 16-QAM : 16 Quadrature Amplitude Modulation, [26](#)
- 3G : Third-Generation, [26](#)
- 3GPP : 3rd Generation Partner Project, [28](#)
- 4G : fourth generation, [26](#)
- 64-QAM : 64 Quadrature Amplitude Modulation, [26](#)
- ACK : Acknowledgement, [31](#)
- AMC : Adaptive Modulation and Coding, [23](#)
- AMPS : Advanced Mobile Phone Service, [25](#)
- AMR : Adaptive Multi Rate, [27](#)
- BE : Best Effort, [35](#)
- BLER : BLock Error Rate, [124](#)
- BR : Bandwidth Request Ranging, [33](#)
- BS : Base Station, [32](#)
- BTS : Base Transeiver Station, [32](#)
- CAC : call admission control, [36](#)
- CBR : Constant Bit Rate, [154](#)
- CDMA : Code Division Multiple Access, [20](#)
- CQI : Channel Quality Indicator, [24](#)
- DECT : Digital Enhanced Cordless Telecommunications, [25](#)
- DL-MAP : DownLink Map, [125](#)
- DS-CDMA : Direct Sequence-CDMA, [20](#)
- EDGE : Enhanced Data rate for GSM Evolution, [25](#)
- FDMA : Frequency Division Multiple Access, [19](#)
- FFT : Fast Fourier Transform, [131](#)
- FFTH : Fast Fair Throughput, [32](#)
- FTP : File Transfer Protocol, [34](#)
- GPRS : General Packet Radio Service, [25](#)
- H-ARQ : Hybrid Automatic Repeat reQuest, [31](#)
- HDR : High Data Rate, [38](#)
- HR : Hand-Over Ranging, [33](#)
- HS-DSCH : High Speed-Downlink Shared Channel, [31](#)
- HSDPA : High-Speed Downlink Packet Access, [26](#)
- HSPA : High Speed Packet Access, [23](#)
- HSUPA : High Speed Uplink Packet Access, [117](#)
- IR : Initial Ranging, [33](#)
- ITT : Interval Time Transmission, [31](#)
- LCH : Lost Call Held, [36](#)
- LTE : Long Term Evolution, [13](#)
- MAC : Medium Access Control layer, [33](#)
- MAN : Metropolitan Area Network, [32](#)
- MCS : Modulation and Coding Scheme, [24](#)
- MDT : Mean Delay Time, [38](#)
- ME : Mobile Equipment, [29](#)
- MMSE : Minimum Mean Square Error, [37](#)
- MR-ACELP : multi-rate ACELP, [27](#)
- NACK : Negative-Acknowledgement, [31](#)
- NLOS : Non Line Of Sight, [32](#)
- NMT : Nordic Mobile Telephony, [25](#)
- NRT : Non-Real-Time, [36](#)
- OFDM : Orthogonal Frequency Division Multiple, [22](#)
- OFDMA : Orthogonal Frequency Division Multiple Access, [21](#)

PF : Proportional Fair, [32](#)

PR : Periodic Ranging, [33](#)

QoS : Quality of Service, [28](#)

QPSK : Quadrature Phase Shift Keying, [26](#)

RNC : Radio Controller Network, [30](#)

RNS : Radio Network Subsystem, [30](#)

RRM : Radio Resource Management, [38](#)

RT : Real-Time, [36](#)

RWP : Random Waypoint, [137](#)

TACS : Total Access Cellular System, [25](#)

TCP : Transmission Control Protocol, [37](#)

TCP : Transport Control Protocol, [38](#)

TDMA : Time Division Multiple Access, [20](#)

UMTS : Universal Mobile Telecommunication Services, [26](#)

USIM : Universal Subscriber Identity Module, [29](#)

W-CDMA : Wideband-Code Division Multiple Access, [20](#)

WiMAX : Worldwide Interoperability for Microwave Access, [23](#)

Bibliographie

- (3GPP, 2000) 3GPP, 2000. Technical Groupe Services and System Aspects; QoS Concepts and Architecture (Release 1999). Tech. Rep. TS 23.107 v3.2.0, <http://www.3gpp.org>.
- (Altman, 2002) E. Altman, 2002. Capacity of multi-service CDMA cellular networks with best-effort applications. Dans les actes de *Dans les actes de MOBICOM*, Atlanta, Georgia, USA.
- (Altman et al., 2004a) E. Altman, K. Avrachenkov, et R. Nunejz-Queija, 2004a. Perturbation analysis for denumerable markov chains with application to queuing models. *36, Number 3*, 839–853.
- (Altman et al., 2004b) E. Altman, K. Avrachenkov, et B. Prabhu, 2004b. A singular perturbation approach to analysing a red queue. Dans les actes de HET-NETs'04 International working conference on Performance Modelling and Evaluation of Heterogeneous Networks.
- (Alwis, 2005) P. M. D. Alwis, 2005. call admission control and resource utilization in wcdma networks. Thesis, Department of Electrical and Computer Engineering University of Canterbury Christchurch, New Zealand.
- (Ameigeiras et al., 2004) P. Ameigeiras, J. Wigard, et P. Mogensen, 2004. Performance of packet scheduling methods with different degree of fairness in hsdpa. Vehicular Technology Conference, VTC-Fall. IEEE 60th Vol. 2, pp. 860-864.
- (Andrews, 2007) J. G. Andrews, 2007. *Fundamentals of WiMAX : Understanding Broad-band Wireless Networking*. ISBN : 0132225522, Prentice Hall PTR, Pages : 496.
- (Assaad, 2006) M. Assaad, 2006. *Cross Layer Study in HSDPA System*. Thèse de Doctorat.
- (Avrachenkov, 1999) K. Avrachenkov, 1999. Analytic perturbation theory and its applications. PhD Thesis, University of South Australia. <http://citeseer.ist.psu.edu/avrachenkov99analytic.html>.
- (Avrachenkov, 2004) K. E. Avrachenkov, 2004. Singularly perturbed finite Markov chains with general ergodic structure. citeseer.ist.psu.edu/722650.html.

- (Avrachenkov et Haviv, 2004) K. E. Avrachenkov et M. Haviv, 2004. The first laurent series coefficients for singularly perturbed stochastic matrices. 386, 243–259.
- (Ball et al., 2005) C. Ball, F. Trend, X. Gaube, et A. Klein, 2005. Performance analysis of temporary removal scheduling applied to mobile WiMAX scenarios in tight frequency reuse. *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*.
- (Baskett et al., 1975) F. Baskett, K. Chandy, R. Muntz, et F. Palacios, 1975. Open, closed and mixed networks of queues with different classes of customers. *Journal of the ACM (JACM)*, Vol. 22(2), pp. 248-260.
- (Benameur et al., 2001) N. Benameur, S. B. Fredj, F. Delcoigne, S. Oueslati-Boulahia, et J.W.Roberts, 2001. Integrated admission control for streaming and elastic traffic. *Dans les actes de QofIS 2001, Coimbra, Septembre*.
- (Bettstetter et al., 2004) C. Bettstetter, H. Hartenstein, et X. P´erez-Costa, 2004. Stochastic properties of the random waypoint mobility model. *ACM/KluwerWireless Networks : Special Issue on Modeling and Analysis of Mobile Networks, 10 (September)*.
- (Bonald et Proutière, 2003) T. Bonald et A. Proutière, 2003. Wireless downlink data channels : user performance and cell dimensioning. *Dans les actes de Dans les actes de MobiCom '03 : the 9th annual international conference on Mobile computing and networking, New York, NY, USA, 339–352. ACM*.
- (Bonald et Proutière, 2005) T. Bonald et A. Proutière, 2005. On the Traffic Capacity of Cellular Data Networks. *Dans les actes de International Teletraffic Congress'05*.
- (Burden et Faires, 1997) R. L. Burden et J. Faires, 1997. *Numericals Analysis*. ISBN-10 0534955339.
- (Chahed et al., 2008) T. Chahed, E. Altman, et S. Elayoubi, 2008. Joint uplink and downlink admission control to both streaming and elastic flows in CDMA/HSDPA systems. *Dans les actes de Journal Performance Evaluation, Volume 65, Issues 11-12, Pages 869-882*.
- (Chakka, 1998) R. Chakka, 1998. Spectral expansion solution for some finite capacity queues.
- (Chakka et Mitrani, 1995a) R. Chakka et I. Mitrani, 1995a. Spectral expansion solution for a class of markov models : Application and comparison with the matrix-geometric method.
- (Chakka et Mitrani, 1995b) R. Chakka et I. Mitrani, 1995b. Spectral expansion solution for a finite capacity multiserver system in a markovian environment.
- (Chao et al., 2001) X. Chao, M. Miyazawa, M. Pinedo, et B. Atkinson, 2001. Queuing networks : Customers, signals and product form solutions. *The Journal of the Operational Research Society*, Vol. 52, No. 5. (May), pp. 600-601.

- (Downey, 2001) A. B. Downey, 2001. The structural cause of file size distributions. *SIGMETRICS Performance Evaluation Review*, Vol. 29(1), pp. 328-329.
- (Elayoubi et al., 2006) E. Elayoubi, T. Chahed, et G. Habuterne, 2006. Mobility-aware admission control schemes in the downlink of third generation wireless systems. *IEEE Transactions on Vehicular Technology*.
- (Elayoubi et Chahed, 2004) S. E. Elayoubi et T. Chahed, 2004. Admission control in the downlink of WCDMA/UMTS. Dans les actes de *EuroNGI Workshop*, 136–151.
- (Elayoubi et al., 2005) S. E. Elayoubi, T. Chahed, et L. Salahaldin, 2005. Optimization of radio resource management in UMTS using pricing. *Computer Communications* 28(15).
- (Elayoubi et al., 2004) S. E. Elayoubi, T. Chahed, M. Tlais, et A. Samhat, 2004. Measurement-based admission control in UMTS. *Special Issue of Annals of Telecommunications on Traffic Engineering and Routing*.
- (Elazouzi et Altman, 2004) R. Elazouzi et E. Altman, 2004. Aqueuing analysis of packet dropping over a wireless link with retransmissions. *Dans les actes de IEEE ICC*.
- (Evans et Everitt, 1999) J. S. Evans et D. Everitt, 1999. On the Teletraffic Capacity of CDMA Cellular Networks. Dans les actes de *IEEE Transactions on Vehicular Technology*, vol. 48.
- (Fodor et Telek, 2005) G. Fodor et M. Telek, 2005. Blocking Probability Bounds in Multi-service CDMA Networks. Dans les actes de *International Teletraffic Congress, ITC 2005, Beijing, China*.
- (Gilhousen et al., 1990) K. Gilhousen, I. M. Jacobs, R. Padovani, A. J., A. J. Viterbi, L. A. Weaver, et C. E. Wheatley, 1990. On Capacity of a Cellular CDMA System. *IEEE Transactions on Vehicular Technology*, vol. 40.
- (Grossglauser et al., 2002) Grossglauser, M., et D. Tse, 2002. Mobility increases the capacity of ad-hoc wireless networks. Dans les actes de *IEEE/ACM Transactions on Networking*, volume 10, pages 477-486.
- (Hegde et Altman, 2003) N. Hegde et E. Altman, 2003. Capacity of multiservice WCDMA networks with variable gos. Dans les actes de *Dans les actes de IEEE Wireless Communications and Networking Conference (WCNC)*.
- (Holma et Toskala, 2001) H. Holma et A. Toskala, 2001. *WCDMA for UMTS : Radio access for third generation mobile communications*. ISBN 2-7464-0370-6.
- (Holma et al., 2007) H. Holma, A. Toskala, K. Ranta-aho, et J. Pirskanen, 2007. Nokia siemens networks high-speed packet access evolution in 3gpp release 7. *IEEE Communications Magazine*.
- (Hong et Rappaport, 1986) D. Hong et S. Rappaport, 1986. Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and

- nonprioritized handoff procedures. *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. VT-35, NO. 3.*
- (Horrich et al.,) S. Horrich, S. E. Elayoubi, et S. B. Jemaa. On the impact of mobility and joint rrm policies on a cooperative WiMAX/hsdpa network. Dans les actes de WCNC 2008.
- (Hyytiä et Virtamo, 2007) E. Hyytiä et J. Virtamo, 2007. Random waypoint mobility model in cellular networks. *Wireless Network, 13 :177-188.*
- (Ibrahimi et al., 2007) K. Ibrahimi, R. El-Azouzi, A. Kobbane, et E. Bouyakhf, 2007. Capacity evaluation of multi-service W-CDMA : A spectral analysis approach. Dans les actes de Sixth IASTED International Conference, csn.
- (IEEE Std 802.16, 2004) IEEE Std 802.16, 2004. IEEE 802.16-2004, air interface for fixed broadband wireless access systems. IEEE Standard for local and Metropolitan Area Networks.
- (IEEE Std 802.16-2001, 2002) IEEE Std 802.16-2001, 2002. Part 16 : Air interface for fixed broadband wireless access systems.
- (IEEE Std 802.16e, 2005) IEEE Std 802.16e, 2005. IEEE 802.16e/d12, air interface for fixed and mobile broadband wireless access systems. IEEE Standard for local and Metropolitan Area Networks.
- (Javier et Mamadou, 2004) S. Javier et T. Mamadou, 2004. Umts (réseaux et télécommunications, 2ème édition.). Hermes - Science.
- (Jeon et Jeong, 2001) W. S. Jeon et D. G. Jeong, 2001. Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink. *IEEE Transactions on Vehicular Technology*, Volume 50, Issue 1, Page(s) :59 - 66.
- (Kelif et Altman, 2004) J. M. Kelif et E. Altman, 2004. admission and gos control in multiservice WCDMA system. Dans les actes de *Dans les actes de ECUMN'2004*, Porto, Portugal.
- (Kwan et al., 2002) R. Kwan, P.Chong, et M. Rinne, 2002. Analysis of the adaptive modulation and coding algorithm with the multicode transmission. *Journal Vehicular Technology Conference*, Vol. 4, pp. 2007-2011.
- (Laiho et Wacker, 2001) J. Laiho et A. Wacker, 2001. Radio network planing process and methods for WCDMA. *Ann. Telecommun.*, vol 56, no. 5-6, pp. 317-31.
- (Lancaster, 1984) P. Lancaster, 1984. *Theory of matrices*. ISBN 0-12-435560-9,.
- (Latouche et Ramaswami, 1999) G. Latouche et V. Ramaswami, 1999. *Introduction to matrix analytic methods in stochastic modeling*. ISBN 0-89871-425-7 (pbk).
- (Leong et Zhuang, 2002) C. W. Leong et W. Zhuang, 2002. Call admission control for voice and data traffic in wireless communications. *Computer Communications*, vol. 25, no. 10, pp.972-979.

- (Li et al., 2003) B. Li, L. Li, B. Li, et X.-R. Cao, 2003. On handoff performance for an integrated voice/data cellular system. *Wirel. Netw.* 9(4), 393–402.
- (Li et Sampalli, 2007) J. Li et S. Sampalli, 2007. Cell mobility based admission control for wireless networks with link adaptation. Dans les actes de ICC.
- (Liu et al., 2005) Q. Liu, S. Zhou, et G. Giannakis, May 2005. Queuing with adaptive modulation and coding over wireless links : Cross-layer analysis and design. Dans les actes de *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, vol.4, NO.3.
- (Mandayam et al., 1997) N. Mandayam, J. Holtzman, et S. Barberis, 1997. Performance and capacity of a voice/data CDMA system with variable bit rate sources. *special Issue on Insight into Mobile Multimedia Communications, Academic Press Inc.*
- (Mehta et al., 2002) N. Mehta, L. Greenstein, T. Willis, et Z. Kotic, 2002. Analysis and results for the orthogonality factor in WCDMA downlinks. Dans les actes de IEEE Vehicular Technology Conference, pp. 100-104.
- (Meyn et Tweedie, 1994) S. P. Meyn et E. R. L. Tweedie, 1994. Computable bounds for geometric convergence rates of markov chains. 4., No. 4., 981–1011.
- (Mitrani, 1995) I. Mitrani, 1995. The spectral expansion solution method for markov processes on lattice strips. In J.H. Dshalalow, editor. *Advances in Queueing Theory, Methods and Open Problems*, CRC Press, Boca Raton, FL, Chapter 13, pp. 337-352.
- (Nakamura et al., 2002) M. Nakamura, Y. Awad, et S. Vadgama, 2002. Adaptive control of link adaptation for high speed downlink packet access (hsdpa) in w-cdma. Dans les actes de 5th International Symposium on Wireless Personal Multimedia Communications, Vol. 2, pp. 382-386.
- (Neuts, 1981) F. Neuts, 1981. *Matrix-geometric solutions in stochastic models : An algorithmic approach*. ISBN 0-486-68342-7 paperback.
- (Norris, 1999) J. R. Norris, 1999. *Markov chains*. Cambridge University Press, isbn = 0-521-68181-3, pages = 237.
- (Peyre et El-Azouzi, 2009) T. Peyre et R. El-Azouzi, 2009. Ieee802.16 multi-class capacity including amc scheme, mobility and qos differentiation for initial and bandwidth request ranging. To appear in WCNC 2009.
- (Peyre et al., 2008) T. Peyre, R. El-Azouzi, et T. Chahed, 2008. QoS differentiation for initial and bandwidth request ranging in IEEE802.16. *Dans les actes de Personal, Indoor and Mobile Radio Communications (PIMRC), Septembre*.
- (Philips et al., 1981) R. Philips, P. V.Kokotovic, et P. FELLOW, 1981. A singular perturbation approach to modeling and control of markov chains. 1087–1094.
- (Pla et Casares-Giner, 2005) V. Pla et V. Casares-Giner, 2005. Analysis of priority channel assignment schemes in mobile cellular communication systems : a spectral theory approach. *Performance Evaluation*, Vol. 59, pp. 199-224.

- (Qiu et al., 2002) R. Qiu, W. Zhu, et Y. Zhang, 2002. Third-generation and beyond (3.5g) wireless networks and its applications. Dans les actes de IEEE International Symposium, Vol. 1 , pp. I-41-I-44.
- (Queija, 2000) R. N. Queija, 2000. *Processing sharing models for integrated-services networks*.
- (Sari et Karam, 1995) Sari et G. Karam, 1995. Transmission techniques for digital terrestrial tv broadcasting. *IEEE Communication Magazine*, Vol. 33, No. 2.
- (Sartori et al., 2007) L. Sartori, S.-E. Elayoubi, B. Fourestié, et Z. Nour, 2007. On the wimax and hsdpa coexistence. Dans les actes de IEEE ICC, pp. 5636-5641.
- (Seo et al., 2004) H. H. Seo, B. H. Ryu, E. S. Hwang, C. H. Cho, et W. L. N, 2004. A study of code partitioning scheme of efficient random access in OFDMA-CDMA ranging subsystem. Dans les actes de *JCCI 2004*, 262.
- (Sipila et al.,) K. Sipila, Z. Honkasalo, J. Laiho-Steffens, et A. Wackr. Estimation of capacity and required transmission power of WCDMA downlink based on a downlink pole equation. Dans les actes de IEEE VTC 2000.
- (Smith et Collins, 2002) C. Smith et D. Collins, 2002. 3g wireless networks. McGraw-Hill Professional.
- (Tang et Goldsmith, 2001) X. Tang et A. Goldsmith, 2001. Admission control and adaptive CDMA for integrated voice and data systems. Dans les actes de IEEE VTC, pp. 506-510, Rhodes, Greece.
- (Tarhini et Chahed, 2007a) C. Tarhini et T. Chahed, 2007a. On capacity of OFDMA-based IEEE802.16 WiMAX including adaptive modulation and coding (AMC) and inter-cell interference. *Dans les actes de Local and Metropolitan Area Networks*.
- (Tarhini et Chahed, 2007b) C. Tarhini et T. Chahed, 2007b. On capacity of OFDMA-based IEEE802.16 WiMAX including adaptive modulation and coding and inter-cell interference. Dans les actes de LANMAN'2007, Princeton NJ, June.
- (Viterbi et al., 1994) A. Viterbi, A. Viterbi, et E. Zehavi, 1994. Other-cell interference in cellular power-controlled CDMA. *IEEE Transactions on Communications*, 42 1501-1504.
- (Viterbi, 1995) A. J. Viterbi, 1995. *CDMA Principles of Spread Spectrum Communication*. Addison-Wesley.
- (Viterbi et Viterbi, 1993) A. M. Viterbi et A. J. Viterbi, 1993. Erlang Capacity of a Power Controlled CDMA System. *IEEE Journal on Selected Areas in Communications*, Vol. 11.
- (Wang et al., 2005) H. Wang, W. Li, et D. P. Agrawal, Avril 2005. Dynamic admission control and QoS for 802.16 wireless MAN. *Wireless Telecommunications Symposium*.

- (Won et al., 2003) J. J. Won, H. H. Seo, B. H. Ryu, C. H. Cho, et H. W. Lee, 2003. Performance analysis of random access protocol in OFDMA-CDMA. Dans les actes de *KICS Fall Conference*.
- (Wongthavarawat et Ganz, 2003) K. Wongthavarawat et A. Ganz, Février 2003. Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems. *International Journal of Communication Systems, Vol. 16, pages 81-96*.
- (Zhang et Yue, 2001) Q. Zhang et O. Yue, 2001. UMTS air interface voice/data capacity-part 1 : reverse link analysis. Dans les actes de *IEEE Veh. Technol. Conf*, 2725–2729.