

Contrôle d'admission sur une cellule IEEE802.16e

Contents

4.1	Formulation du problème	79
4.1.1	Objectifs	79
4.1.2	Environnement	80
4.1.3	Description du système	81
4.1.4	Algorithme du contrôle d'admission	81
4.2	Modèle théorique	82
4.2.1	Définition des états du système	82
4.2.2	Arrivées et départs du système	84
4.2.3	Transitions d'état du système	87
4.3	Mesures de performance	90
4.3.1	Débit total des trafics non temps réel	90
4.3.2	Temps moyen de séjour des appels non temps réel	91
4.3.3	Probabilité de perte des appels temps réel	91
4.3.4	Probabilité de blocage des trafics temps réel	92
4.4	Résultats et analyses numériques	93
4.4.1	Performances sans mobilité des utilisateurs	94
4.4.2	Performance en environnement mobile	98
4.5	Bilan et perspectives	102

4.1 Formulation du problème

4.1.1 Objectifs

Le chapitre précédent identifie et modélise les performances sur couche MAC des utilisateurs IEEE802.16e. En outre, nous y avons intégré un principe nouveau de par-

titionnement logique des codes CDMA entre les deux principaux types de trafic. La continuité de cette étude vise à faire de même au niveau structurel supérieur. En effet, nous désirons maintenant étudier la capacité d'une cellule IEEE802.16e à accueillir des utilisateurs. L'objectif est alors de modéliser et d'étudier le comportement d'un tel système en fonctionnement. Un autre objectif est de proposer différents mécanismes liés au contrôle d'admission. Dans ce chapitre, nous cherchons à étudier les performances atteintes par notre algorithme de contrôle d'admission. Cet algorithme se définit au travers de ses objectifs. Premièrement, il doit fournir aux trafics temps réel les ressources demandées. Deuxièmement, il doit assurer accès aux ressources égalitaire entre les trafics non temps réel. En outre, l'algorithme de CAC doit aussi empêcher toute famine de ressources pour ces même appels. Enfin, il doit maintenir la plus grande efficacité possible vis à vis de l'utilisation des ressources radio. Par ailleurs, nous intégrons dans ce modèle la mobilité intra-cellulaire des utilisateurs. Cette intégration nouvelle introduit le besoin d'adjoindre à notre algorithme de CAC des éléments lui permettant de gérer au mieux cette nouvelle donnée. Nous désirons donc aussi étudier la capacité d'une cellule IEEE802.16e à assurer une pérennité de service à des utilisateurs mobiles. L'objectif est alors de modéliser et d'étudier le comportement d'un tel système en fonctionnement. Un autre objectif est de proposer une autre aptitude liées au contrôle d'admission. Ce nouvel algorithme devra limiter la perte des appels lorsque les utilisateurs se déplacent d'un secteur à un autre (voir la section 1.2.3 pour plus de détails).

4.1.2 Environnement

Aussi, nous développons dans ce chapitre un modèle de capacité pour une cellule IEEE802.16e. Ce modèle intègre le mécanismes de différenciation de trafic du chapitre précédent. En se basant sur le processus général des arrivées de requêtes, nous proposons ici un nouvel algorithme d'allocation de ressources pour les trafics temps réel basé sur la position de l'utilisateur (SNR) dans la cellule.

Dans ce système, toutes les connexions temps réel disposent du même débit et les durées de ces appels sont indépendantes de la quantité de ressources consommées. En effet, l'expérience et la logique prouve que le temps de séjour d'un appel temps réel dans un tel système est uniquement lié aux habitudes comportementales des utilisateurs vis à vis de la consommation de ce genre de service. A l'inverse, la durée des appels non temps réel dépendent de l'assignation dynamique des ressources réparties parmi tous les appels de ce type.

En ce qui concerne la mobilité des utilisateurs, et donc de leur appels, ils peuvent désormais changer de secteur de modulation au cours du temps (voir section 1.2.3). Le comportement de migration ne dépend pas de la classe de trafic. Les appels sont libres de migrer uniquement vers les secteurs adjacents et à raison d'un maximum de une migration par appel et par temps du système.

4.1.3 Description du système

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, ces mêmes études supposent l'indépendance des différents types de connexions engagées dans le standard IEEE802.16e. Or, ces connexions utilisent un nombre fini de codes soumis à la compétition des mobiles. Par conséquent, les différents processus d'arrivée sont alors implicitement dépendants. Cet état de fait est un des points de conclusion du chapitre précédent (voir section 3.4).

Dans ce chapitre, nous développons le modèle d'une seule cellule IEEE802.16e composée de deux secteurs utilisant deux modulations différentes (voir la section 1.2.3 pour plus de détails). Ces appels peuvent être des appels temps réel (RT), correspondant aux trafics de type UGS et rtPS, ou non temps réel (NRT) pour les trafics nrtPS et BE (voir section 1.3.5). L'accès aux ressources de la cellule se fait en accord avec le principe de différenciation des plages de codes accessibles aux trafics temps réel et non temps réel, tel que défini dans le chapitre précédent. Tous les mobiles de la cellule suivent un comportement saturé : ils disposent continuellement d'un service en attente de ressources radio.

La cellule est composée de r secteurs. Nous considérons que les mobiles sont uniformément distribués sur l'ensemble de la cellule. Ainsi, la population de chaque secteur est déterminée en fonction de l'aire de chacun d'eux. Enfin, la modélisation se base sur une chaîne de Markov à temps discret (DTMC), où un temps système correspond à la durée d'une trame MAC.

4.1.4 Algorithme du contrôle d'admission

Nous définissons maintenant l'algorithme de contrôle d'admission (CAC) des requêtes entrantes dans le système à temps discret.

Les trafics temps réel sont caractérisés par un même débit binaire. En conséquence, ces trafics reçoivent un nombre de sous-porteuses (voir section 1.2.1 et figure 1.2) correspondant à l'efficacité de leur modulations (bit par symbole), et ainsi en fonction de leur secteurs respectifs. 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 (*batch arrivals*). Dans le cas où le système ne peut pas satisfaire toutes ces demandes à la fois, l'algorithme de contrôle d'admission acceptera un certain nombre d'entre elles en favorisant celles qui demandent le moins de ressources. Par conséquent, ce seront celles issues des secteurs les plus intérieurs à la cellule qui seront satisfaites en priorité. Ce principe a pour principal objectif d'assurer une gestion efficace de la ressource radio. L'acceptation prioritaire des meilleures modulations permet de maximiser la quantité globale d'information transmise par sous-porteuse. 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.

A l'inverse, les trafics non temps réel n'ont aucune exigence précise de bande passante. Le contrôle d'admission allouera ici un nombre de sous-porteuses identiques à chacun de ces trafics. Les appels non temps réel supportent la variation dynamique de leur ressources allouées. Aussi, à chaque temps du système, la station de base leur allouera les ressources laissées disponibles par les appels temps réel en cours. Le partage de ces ressources se fait sur la base d'un processus égalitaire de (*Processor Sharing*) (Benameur et al., 2001). Ainsi, le débit final d'un trafic non temps réel dépend à la fois de son secteur (c'est à dire de sa modulation), des ressources laissées libres 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 ressources 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 les performances de cet algorithme en fonction de cette quantité réservée. Par cette précaution, les trafics non temps réel ne pourront jamais être bloqués à l'entrée du système. Rappelons, enfin, que le temps de séjour de ces appels dépend intrinsèquement des ressources consommées : plus le trafic disposera de sous-porteuse, plus vite le service se terminera.

Par ailleurs, notre algorithme de contrôle d'admission vise à réduire les probabilités qu'un appel temps réel soit perdu d'un secteur à un autre faute de ressources encore libres : si un utilisateur appelant migre d'un secteur vers un autre plus en bordure de cellule, l'appel demandera un supplément de sous-porteuses afin de compenser la perte d'efficacité de modulation induite par le changement de secteur. Dans la mesure où les ressources encore disponibles dans le système sont insuffisantes, l'appel sera perdu et l'utilisateur extrêmement mécontent. Afin d'éviter ces pertes, le CAC proposé ici introduit une autre réserve de bande passante. Cette réserve vise à satisfaire le besoin de supplément de bande passante des appels migrant en bordure de cellule. Quand plusieurs migrations de ce type interviennent durant un même temps système, le CAC acceptera en priorité les migrations demandant le plus faible supplément de ressources, c'est à dire celles intervenant dans les secteurs les plus intérieurs de la cellule. En conséquence, du moment qu'un appel est perdu dans un secteur après migration, toutes les migrations dans les dans un autre secteur, extérieurs à celui-ci seront aussi perdues.

4.2 Modèle théorique

4.2.1 Définition des états du système

Le système gère deux classes de service sur r secteurs. 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 secteurs.

Soit n_i^k le nombre d'appels résidant dans le système pour la classe k , $k = \{RT, NRT\}$

et dans le secteur $i, i = 1, \dots, r$. Pour les besoins de l'étude, et pour simplifier les écritures, nous différencions les vecteurs \vec{n}^{RT} et \vec{n}^{NRT} représentant respectivement le nombre des appels RT et NRT du système. Le vecteur \vec{n} est défini de la manière suivante :

$$\vec{n} = (\vec{n}^{RT}, \vec{n}^{NRT}) \quad (4.1)$$

$$\vec{n} = (n_1^{RT}, \dots, n_r^{RT}, n_1^{NRT}, \dots, n_r^{NRT}) \quad \vec{n} \in \mathbb{N}^{2r} \quad (4.2)$$

Pour la suite, soit L la bande passante totale du système. Comme décrit dans l'algorithme du contrôle d'admission, nous établissons L_{min}^{NRT} la bande passante minimale réservée aux appels NRT. Soit L_{mob} une autre partie de bande passante réservée à la gestion de la mobilité des utilisateurs. Naturellement, les appels RT auront un accès prioritaire à la bande passante laissée après ces deux parties réservées : L^{RT}

$$L^{RT} = L - L_{min}^{NRT} - L_{mob} \quad (4.3)$$

Aux appels RT sont assignées un certain nombre de sous-porteuses par secteur L_i^{RT} pris parmi L^{RT} . Ce nombre de sous-porteuses dépend du débit donné à chaque appel RT dans la cellule et de la modulation utilisée par le mobile engageant ce type d'appel. Dès lors, les appels NRT se partagent la bande passante restante L^{NRT} . Par contre, cette dernière ressource sera partagée entre les appels NRT en suivant le principe défini dans (Benameur et al., 2001). Ce partage est défini aux travers des relations suivantes :

$$L^{NRT}(\vec{n}^{RT}) = L - \sum_{i=1}^r n_i^{RT} L_i^{RT} \quad (4.4)$$

$$\sum_{i=1}^r n_i^{RT} L_i^{RT} \leq L^{RT} \quad (4.5)$$

$$L_i^{NRT}(\vec{n}) = \frac{L^{NRT}(\vec{n}^{RT})}{\sum_{i=1}^r n_i^{NRT}} \quad (4.6)$$

Nous précisons que la variable $L_i^{NRT}(\vec{n})$ accepte de prendre des valeurs non entières. En effet, l'allocation des ressources peut se faire suivant un multiplexage temporel. Cette particularité autorise donc qu'un appel non temps réel reçoive un nombre décimal de sous-porteuses une fois rapporté à la proportion d'un seul *slot* système.

A partir des relations définies dans (Tarhini et Chahed, 2007b), nous calculons le débit physique. Notons que le débit d'un appel temps réel (RT) est fixé par les informations contenues dans la requête de bande passante à l'origine de cet appel. Nous supposons que tous les appels temps réel demandent au système le même débit de transmission. Soit R^{RT} le débit requis par chaque appel temps réel. Soit R_i^{NRT} le débit atteint par les appels non temps réel dans la région i . Ce débit répond aux caractéristiques particulières décrites précédemment.

Afin de procéder au calcul du débit des appels NRT, nous définissons K comme le nombre de fréquences assignées à chaque sous-porteuse de transmission de données

temps réel. Soit B le débit symbole (nombre de symboles par seconde) et E_i l'efficacité de modulation (en nombre de bits par symbole). $BLER_i$ correspond au taux d'erreur bloc (BLER) d'une région i . De plus, nous définissons L_i^k comme le nombre de porteuses assignées à un appel de classe k dans le secteur i . Le débit R_i^k pour un appel de classe k dans le secteur i est alors donné par la relation suivante :

$$R_i^k(\vec{n}) = L_i^k(\vec{n}) \times K \times B \times E_i \times (1 - BLER_i) \quad (4.7)$$

Par ailleurs, nous pouvons aisément déterminer le nombre de sous-porteuses requises par un appel RT dans le secteur i :

$$L_i^{RT} = \frac{R^{RT}}{K \times B \times E_i \times (1 - BLER_i)} \quad (4.8)$$

L'espace des 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 :

$$E = \{ \vec{n} \in \mathbb{N}^{2r} \mid \sum_{i=1}^r n_i^{RT} L_i^{RT} \leq L^{RT} \} \quad (4.9)$$

4.2.2 Arrivées et départs du système

Nous caractérisons ici les arrivées et départs d'appel qui interviennent dans le système. Les arrivées peuvent être la conséquence d'une migration provenant des secteurs voisins ainsi que l'acceptation dans le système d'une nouvelle requête initialisant un nouvel appel. De la même manière, un départ d'appel peut être le résultat d'une migration vers un secteur voisin ou alors de la fin d'un appel.

Premièrement, nous calculons le processus d'arrivée par engagement des nouveaux appels entrant dans le système.

Sachant que les requêtes de type RT et NRT proviennent de chacun des r secteurs, nous considérons que les n mobiles sont uniformément répartis sur la cellule IEEE802.16e. Aussi, le nombre n_i de mobiles contenu par le secteur i est fonction de son ratio de superficie (voir table 1.2). Soit r_i le rayon du secteur i avec $r_0 = 0$ et R le rayon total de la cellule. La population n_i d'un secteur i est donnée par la relation suivante :

$$n_i = \frac{r_i^2 - r_{i-1}^2}{R^2} \times n \quad \text{avec } i \in [1, \dots, r] \quad (4.10)$$

Soit \vec{A} et \vec{B} , deux vecteurs composés de r variables aléatoires à valeurs entières indiquant les nouvelles arrivées dans chacun des r secteurs. \vec{A} pour les trafics RT et \vec{B} pour les trafics NRT. Soit \vec{Z} le vecteur de dimension $2r$ représentant les nouveaux appels arrivant sur tous les secteurs pour les deux types de trafics :

$$\vec{A} = (A_1, \dots, A_r), \quad \vec{B} = (B_1, \dots, B_r), \quad \text{et } \vec{Z} = (\vec{A}, \vec{B}) \quad (4.11)$$

Nous considérons que les arrivées pour les appels RT et NRT sont dépendantes. Le nombre maximum d'appels RT (respectivement NRT) est fonction du nombre de codes associés à ce type de trafic : $N_{RT} = N_1 + N_2$ (respectivement $N_{NRT} = N_2$). Remarquons aussi que le nombre total des arrivées ne peut pas excéder le nombre total des codes N . Suivant le principe de partitionnement en vigueur ici : $A = \sum_i A_i$, $B = \sum_i B_i$ et $A + B \leq N_{RT}$.

Pour la suite de l'étude, nous identifions $P(\vec{Z})$, la probabilité associée au vecteur \vec{Z} . Cette probabilité suit la relation suivante :

$$P(A_1 = a_1, \dots, B_r = b_r) = \frac{\prod_{i=1}^r \binom{n_i}{a_i} \binom{n_i - a_i}{b_i}}{\binom{n}{a} \binom{n - a}{b}} P(A = a, B = b) \quad (4.12)$$

où

$$a = \sum_{i=1}^r a_i \quad \text{et} \quad b = \sum_{i=1}^r b_i \quad (4.13)$$

Notons que la probabilité $P(A = a, B = b)$ correspond à la distribution jointe donnée par notre simulateur comportemental. Cette distribution peut d'ailleurs être approximé par la méthode décrite dans la section 3.3.4.

Deuxièmement, nous calculons les distributions relatives aux processus de migration des appels. Soit $p_{i,j}$ la probabilité qu'un appel dans le secteur i migre vers le secteur j . Afin de calculer cette probabilité pour chaque classe de trafic dans chaque secteur, nous définissons préalablement $m_{i,j}^k$ comme la variable aléatoire du nombre d'appels de classe k dans le secteur i migrant vers la région j . $m_{i,j}^k \in [0, n_i^k]$. Notons qu'ici, nous devons uniquement prendre en compte les appels toujours en cours à la fin du temps système en omettant ceux qui auront terminés leur services durant cet espace de temps. Aussi, la distribution des migrations intervenant dans la cellule se base en partie sur la connaissance de la répartition des appels finissant leur services durant l'espace de temps concerné. En se basant sur ces observations, la probabilité de migration est définie de la manière suivante :

$$P(m_{i,j}^k = x | D^c = y) = \binom{n_i^k - y}{x} (p_{i,j})^x (1 - p_{i,j})^{n_i^k - x - y} \quad (4.14)$$

Enfin, nous désirons déterminer le processus de départ des appels du système suite à la terminaison de leur services. Nous calculons ainsi les probabilités de départ des appels de classe k dans un secteur i . Sachant que la durée des appels RT ne dépend pas des ressources consommées. Les appels RT demeurent dans le système durant un temps exponentiellement distribué de moyenne μ^{RT} . Cette moyenne ne varie pas au cours 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 et le secteur de l'appel considéré. A chaque temps système, le nombre de services en cours évolue, ainsi que les ressources laissées libres par les appels RT. Autre particularité à soulever ici, 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 leur ressources allouées uniquement en début de trame MAC via la DL/UL-MAP (voir section 1.2.1 pour plus de détails). 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 le secteur k est fonction de la taille moyenne de l'entité de données à transmettre $E(Pay)$ et du débit de cet appel R_i^{NRT} . Ce temps étant défini comme suit :

$$\mu_i^{NRT}(\vec{n}) = \frac{R_i^{NRT}(\vec{n})}{E(Pay)} \quad (4.15)$$

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

$$\lambda_i^{RT} = n_i^{RT} \mu_i^{RT} T \quad \text{et} \quad \lambda_i^{NRT} = n_i^{NRT} \mu_i^{NRT}(\vec{n}) T \quad (4.16)$$

Enfin, nous déterminons ici la distribution des fins de service en supposant qu'elles suivent une distribution de Poisson. Soit S_i^k le nombre de services terminés durant un temps système. leur distributions suivent l'expression suivante :

$$P(S_i^k = x | \vec{n}) = \begin{cases} \frac{(\lambda_i^{RT})^x}{x!} e^{-\lambda_i^{RT}}, & \text{si } k = RT \\ \frac{(\lambda_i^{NRT})^x}{x!} e^{-\lambda_i^{NRT}}, & \text{si } k = NRT. \end{cases} \quad (4.17)$$

Finalement, nous pouvons calculer la distribution des fins d'appel intervenant durant une trame MAC. Soit D_i^k le nombre des appels de classe k dans le secteur i , terminant leur services durant une trame MAC :

$$P(D_i^k = d_i^k) = \frac{P(S_i^k = d_i^k)}{\sum_{j=0}^{n_i^k} P(S_i^k = j)} \quad (4.18)$$

4.2.3 Transitions d'état du système

A partir de là, 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 arrivées et départs ayant eu lieu durant tout l'espace de temps de la trame t . Notons aussi que l'ensemble des vecteurs \vec{n}' appartiennent aussi à l'espace E défini par la relation (4.9) :

$$\vec{n}' = (n_1'^{RT}, \dots, n_r'^{RT}, n_1'^{NRT}, \dots, n_r'^{NRT}) \quad \text{avec} \quad \vec{n}' \in E \quad (4.19)$$

Par conséquent, les probabilités de transitions du système sont obtenues en considérant toutes les transitions possibles 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}' . Il correspond aux 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 :

$$P(\vec{n}, \vec{n}') = P(\vec{n}' = \vec{n} + x) \quad (4.20)$$

Les calculs de probabilités de transitions demandent donc de considérer les transitions possibles pour chaque trafic et secteur. Mais, remarquons que ces évolutions dépendent des ressources disponibles L^{RT} . En effet, le système supporte un nombre limité d'appels temps réel défini à travers la relation (4.9). Du fait des principes directeurs de notre algorithme de contrôle d'admission, le système acceptera prioritairement les appels RT exploitant la modulation la moins gourmande en sous-porteuses. De leur côté, les appels non temps réel sont acceptés sans limites et indépendamment des appels RT déjà acceptés. Nous précisons que notre approche discrète des événements considère tout d'abord les départs d'appel, puis les migrations et enfin les nouveaux appels entrant dans le système. L'étude nécessite alors d'introduire les cas où l'occupation en ressource des appels RT atteint les limites du système. Dans ces cas là, nous définissons l'existence de deux éléments i^* et i^\sharp . Le premier élément existe dès lors qu'un appel entrant est bloqué alors que toutes les requêtes envoyées durant la trame MAC n'ont pas encore été traitées. Ainsi, soit i^* , $i^* = 1, \dots, r$, le premier secteur où au moins une requête est bloquée faute de ressources disponibles à sa satisfaction. Le second élément existe lorsqu'une perte d'un appel migrant survient au sein de la cellule. Aussi, soit i^\sharp , $i^\sharp = 1, \dots, r$, le premier secteur où au moins une migration entrante a été bloquée, là aussi faute de ressources disponibles.

Dès lors, nous définissons les comportements de transition propre à chaque cas :

- les transitions du système définissent l'existence d'un secteur spécifique i^\sharp , $1 \leq i^\sharp \leq r$, au delà duquel aucune nouvelle migration vers une région extérieure ne pourra être acceptée. Les appels migrant depuis le secteur i^\sharp vers le secteur $i^\sharp + 1$ sont tous perdus. Ceux migrant du secteur $i^\sharp - 1$ vers i^\sharp sont acceptés jusqu'à la limite $x_{i^\sharp}^{RT}$. Enfin, toutes les migrations arrivant aux secteurs $j = 1, \dots, i^\sharp - 1$ sont naturellement acceptées.

- les transitions du système définissent l'existence d'un secteur spécifique i^* , $1 \leq i^* \leq r$, où tous les appels arrivant des secteurs $j = 1, \dots, i^* - 1$ sont acceptés. Les appels arrivant en secteur i^* sont acceptés jusqu'à la limite $x_{i^*}^{RT}$. Enfin, les appels entrant dans les secteurs $j = i^* + 1, \dots, r$ sont logiquement bloqués.
- Les événements intervenant durant une trame MAC ne permettent à aucun instant d'atteindre les limites du système. Aussi, toutes les migrations ainsi que toutes les arrivées des nouveaux appels sont acceptées et aucun élément i^* ou i^\sharp n'existent.

Les valeurs de i^* et i^\sharp sont données par les relations suivantes :

$$i^* = \min \left(i | L^{RT} - \sum_{j=1}^r L_j^{RT} (n_j^{RT} - d_j^{RT}) \right) \quad (4.21)$$

$$+ \sum_{j=2}^r m_{j,j-1} (L_j - L_{j-1}) \quad (4.22)$$

$$- \sum_{j=1}^{r-1} m_{j,j+1} (L_{j+1} - L_j) \quad (4.23)$$

$$- \sum_{j=1}^i a_j L_j^{RT} < 0) \quad (4.24)$$

$$i^\sharp = \min \left(i | L - L_{min}^{NRT} - \sum_{j=1}^r L_j^{RT} (n_j^{RT} - d_j^{RT}) \right) \quad (4.25)$$

$$+ \sum_{j=2}^r m_{j,j-1} (L_j - L_{j-1}) \quad (4.26)$$

$$- \sum_{j=1}^i m_{j,j+1} (L_{j+1} - L_j) < 0) \quad (4.27)$$

avec $1 \leq i^* \leq r$ et $1 \leq i^\sharp \leq r$

Ici, nous introduisons la fonction δ comme la fonction indicatrice définie comme suit :

$$\delta(X) = \begin{cases} 1, & \text{si } X \text{ est vrai} \\ 0, & \text{sinon.} \end{cases} \quad (4.28)$$

Du moment que i^* existe, nous définissons les fonctions suivantes qui décrivent les événements possibles aboutissant à une surcharge du système. A partir du secteur i^* un appel ou plus est bloqué faute de ressources disponibles.

Afin de faciliter la lisibilité des fonctions de Dirac à venir, nous définissons une relation intermédiaire y_j^k représentant la variation du système pour la classe k dans le

secteur j où les variations dues aux migrations n'ont pas été prises en compte :

$$y_j^k = x_j^k - m_{j-1,j}^k - m_{j+1,j}^k + m_{j,j-1}^k + m_{j,j+1}^k \quad (4.29)$$

$$\delta_1^* = \prod_{j=1}^{i^*-1} \delta(y_j^{RT} = a_j^{RT} - d_j^{RT}) \quad (4.30)$$

$$\delta_2^* = \delta(-x_{i^*}^{RT} - m_{i^*,i^*-1}^{RT} - m_{i^*,i^*+1}^{RT} \leq d_{i^*}^{RT} < a_{i^*}^{RT} - y_{i^*}^{RT}) \quad (4.31)$$

$$\delta_3^* = \prod_{j=i^*+1}^r \delta(2x_j^{RT} - y_j^{RT} = d_j^{RT}) \quad (4.32)$$

Du moment que i^\sharp existe, nous définissons les fonctions suivantes qui aboutissent là aussi à une surcharge du système. A partir du secteur i^\sharp , une migration ou plus venant du secteur inférieur est perdue.

$$\delta_1^\sharp = \prod_{j=1}^{i^\sharp-1} \delta(y_j^{RT} = d_j^{RT}) \quad (4.33)$$

$$\delta_2^\sharp = \delta(-x_{i^\sharp}^{RT} - m_{i^\sharp,i^\sharp-1}^{RT} - m_{i^\sharp,i^\sharp+1}^{RT} \leq d_{i^\sharp}^{RT} < -y_{i^\sharp}^{RT}) \quad (4.34)$$

$$\delta_3^\sharp = \prod_{j=i^\sharp+1}^r \delta(x_j^{RT} = d_j^{RT} + m_{j,j+1}^{RT} + m_{j,j-1}^{RT} - m_{j+1,j}^{RT}) \quad (4.35)$$

Enfin, notons que si i^* et i^\sharp existent en même temps, le calcul de transition requiert seulement l'utilisation des fonctions δ^\sharp . En effet, du moment que i^\sharp existe, aucun nouvel appel RT n'est accepté par le système. Ce choix découle de l'observation qu'il est beaucoup plus pénalisable de perdre un utilisateur déjà consommateur de service par rapport à bloquer un autre demandeur d'un engagement de service.

Par conséquent, les probabilités de migration sont obtenues en calculant tous les cas possibles de départ, d'arrivée et de migration durant une unique trame MAC au travers de la fonction Ω . Ensuite, nous observons quel scénario parmi ceux définis ci-dessus correspond à la transition recherchée via le système Δ .

Le calcul de transition d'état obéit alors aux relations suivantes :

$$P(\vec{n}, \vec{n}') = \sum_{a_1=0}^{s_1} \sum_{b_1=0}^{w_1} \sum_{d_1=0}^{n_1^{RT}} \sum_{y_1^-=0}^{n_1^{RT}} \sum_{y_1^+=0}^{n_1^{RT}} \sum_{z_1^-=0}^{n_1^{NRT}} \sum_{z_1^+=0}^{n_1^{NRT}} \Omega \times \Delta \quad (4.36)$$

avec

$$\begin{aligned}
 f_i &= \max(0, d_i^{RT} + x_i^{RT} - n_{i-1}^{RT} - n_{i+1}^{RT}) \\
 g_i &= \min(N_{RT}, n_i) \\
 v_i &= \max(0, d_i^{NRT} + x_i^{NRT} - n_{i-1}^{NRT} - n_{i+1}^{NRT}) \\
 w_i &= \min(N_{NRT}, n_i) \\
 \Omega &= Z \prod_{i=1}^r (D \cdot M^{RT} \cdot M^{NRT}) \\
 \Delta &= \begin{cases} \delta_1^\# \times \delta_2^\# \times \delta_3^\#, & \text{si } i^\# \in [1, \dots, r]. \\ \delta_1^* \times \delta_2^* \times \delta_3^*, & \text{si } i^* \in [1, \dots, r]. \\ 1, & \text{sinon} \end{cases} \\
 Z &= P(A_1 = a_1, \dots, A_r = a_r, B_1 = b_1, \dots, B_r = b_r) \\
 D &= P(D_i^{RT} = d_i^{RT}) \times \\
 &\quad P(D_i^{NRT} = b_i - x_i^{NRT} - m_{i,i-1}^{NRT} - m_{i,i+1}^{NRT}) \\
 M^{RT} &= P(m_{i,i-1}^{RT} = y_i^- | D^c = d_i^{RT} + y_i^+) \times \\
 &\quad P(m_{i,i+1}^{RT} = y_i^+ | D^c = d_i^{RT} + y_i^-) \\
 M^{NRT} &= P(m_{i,i-1}^{NRT} = z_i^- | D^c = d_i^{NRT} + z_i^+) \times \\
 &\quad P(m_{i,i+1}^{NRT} = z_i^+ | D^c = d_i^{NRT} + z_i^-)
 \end{aligned}$$

En se basant sur toutes ces observations, nous pouvons procéder à la réalisation de la matrice de transition P du système :

$$P = \left(P(\vec{n}, \vec{n}') \right), \text{ pour } (\vec{n}, \vec{n}') \in E \times E \quad (4.37)$$

Dès lors, nous déterminons le vecteur de distribution stationnaire $\vec{\Pi}$ en résolvant le système d'équations indépendant linéaire suivant :

$$\vec{\Pi} = \{ \pi(\vec{n}) | \vec{n} \in E \} \quad \text{avec} \quad \begin{cases} \vec{\Pi} P = \vec{\Pi} \\ \sum_{\vec{n} \in E} \pi(\vec{n}) = 1 \end{cases} \quad (4.38)$$

4.3 Mesures de performance

4.3.1 Débit total des trafics non temps réel

Etant donné que le débit de chaque appel temps réel est fixé par les utilisateurs, nous traitons ici celui des appels non temps réel. Dès lors, nous calculons le débit physique atteint en moyenne par un trafic NRT. Pour cela, nous nous basons sur la relation (4.7), ainsi que le débit moyen à l'état stationnaire du système. Aussi, soit R_{NRT}^{tot} le débit des appels non temps réel de l'ensemble de la cellule.

$$E(R_{NRT}^{tot}) = \sum_{\vec{n} \in E} \pi(\vec{n}) \sum_{i=1}^r n_i^{NRT} R_i^{NRT} \quad (4.39)$$

4.3.2 Temps moyen de séjour des appels non temps réel

Du moment que les appels non temps réel ne sont pas bloqués par le système, ils se partagent les ressources laissées disponibles par les appels temps réel. Or, le temps de séjour des appels non temps réel dépend directement des ressources du système auxquelles ils auront accès. Nous calculons ici le temps moyen de transfert en divisant le nombre moyen d'appels NRT dans un secteur i : $E(NRT_i)$; avec les arrivées moyennes de requêtes de type NRT dans le secteur i : Λ_i^{NRT} .

Le temps moyen de transfert des appels non temps réel dans le secteur i est donné par la loi de Little ([Gelenbe et Pujolle, 1987](#)) :

$$T_i^{NRT} = \frac{E(n_i^{NRT})}{\Lambda_i^{NRT}} = \frac{\sum_{\vec{n} \in E} \pi(\vec{n}) n_i^{NRT}}{\sum_{j=0}^{\min(n_i, N_{NRT})} j P(NRT_i = j)} \quad (4.40)$$

où NRT_i est la variable aléatoire du nombre d'arrivées d'appel non temps réel en région i , dont la probabilité nous est donnée par :

$$P(NRT_i = j) = \sum_{a_1=0}^{c_1, \dots, c_r} \sum_{b_1=0}^{e_1, \dots, e_r} P(a_1, \dots, a_r, b_1, \dots, b_r) \quad (4.41)$$

$$\begin{array}{c} \vdots \\ a_r=0 \\ \vdots \\ b_r=0 \end{array}$$

Pour $l = 1, \dots, r$. $c_l = \min(n_l, N_{RT})$ et $e_l = \min(n_l, N_{NRT})$. $b_i = j$ et $e_i = j$.

4.3.3 Probabilité de perte des appels temps réel

Soit P_D^j la probabilité qu'un appel migrant depuis le secteur j vers le secteur $j + 1$ soit perdu faute de ressources disponibles.

le calcul de cette probabilité requiert le calcul de tous les états et transitions possibles amenant à un dépassement des ressources de la cellule par suite d'une migration ou plus depuis les secteurs 1 à j . Dans ce but, nous avons à calculer la probabilité $p_{i^\#}^j$ qu'un appel temps réel migrant vers le secteur $j + 1$ soit perdu dans la mesure où le secteur bloquant les migrations est identifié : $i^\#$.

La probabilité de perte générale pour chaque secteur est alors définie ainsi :

$$P_D^j = \sum_{\vec{n} \in E} \pi(\vec{n}) \sum_{d_1=0}^{n_1^{RT_1}} \sum_{y_1=0}^{n_1^{RT}} \sum_{y_1^+=0}^{n_1^{RT}} \Omega_D p_{i^\#}^j \quad (4.42)$$

$$y_j^+ = 1 \quad (4.43)$$

$$\Omega_D = \prod_{i=1}^r (D \times M^{RT} \times \delta_1^\# \cdot \delta_2^\# \cdot \delta_3^\#) \quad (4.44)$$

$$D = P(D_i^{RT} = d_i^{RT}) \quad (4.45)$$

$$M^{RT} = P(m_{i,i-1}^{RT} = y_i^- | D^c = d_i^{RT} + y_i^+) P(m_{i,i+1}^{RT} = y_i^+ | D^c = d_i^{RT} + y_i^-) \quad (4.46)$$

$$p_{i^\#}^j = \begin{cases} 0, & \text{si } j < i^\#; \\ 1, & \text{si } j > i^\#; \\ 1 - \frac{\bar{m}_j}{m_j}, & \text{si } j = i^\#. \end{cases} \quad (4.47)$$

où pour $i = 1, \dots, r$:

$$\bar{m}_j = \max\left(m_j | L - L_{min}^{NRT} - \sum_{l=1}^r L_l^{RT} (n_l^{RT} - d_l^{RT}) \right) \quad (4.48)$$

$$+ \sum_{l=2}^r y_l^- (L_l^{RT} - L_{l-1}^{RT}) \quad (4.49)$$

$$- \sum_{l=1}^{i^\#-1} y_l^+ (L_{l+1}^{RT} - L_l^{RT}) \quad (4.50)$$

$$- m_i (L_{i+1}^{RT} - L_i^{RT}) > 0 \quad (4.51)$$

4.3.4 Probabilité de blocage des trafics temps réel

La probabilité de blocage consiste en la probabilité qu'un appel temps réel soit bloqué en entrée du secteur j faute de ressources disponibles lors du traitement de la requête de bande passante.

Soit P_B^j la probabilité qu'un appel désirant entrer en secteur j soit bloqué. Afin de l'obtenir, nous calculons les différentes évolutions d'état du système aboutissant à ces blocages. En outre, nous établissons la probabilité $p_{i^*}^j$ qu'un appel temps réel arrivant au secteur j soit bloqué par le système, pour un i^* donné.

Aussi, la probabilité de blocage est donnée par la relation suivante :

$$P_B^j = \sum_{\vec{n} \in E} \pi(\vec{n}) \sum_{a_1=0}^{g_1} \sum_{b_1=0}^{w_1} \sum_{d_1=0}^{n_{RT_1}} \sum_{y_1^-=0}^{n_1^{RT}} \sum_{y_1^+=0}^{n_1^{RT}} \Omega_B p_{i^*}^j \quad (4.52)$$

$$a_j = 1 \quad (4.53)$$

$$g_i = \min(N_{RT}, n_i) \quad (4.54)$$

$$w_i = \min(N_{NRT}, n_i) \quad (4.55)$$

$$\Omega_B = Z \prod_{i=1}^r (D \times M^{RT} \times \delta_1^* \cdot \delta_2^* \cdot \delta_3^*) \quad (4.56)$$

$$Z = P(A_1 = a_1, \dots, A_r = a_r, B_1 = b_1, \dots, B_r = b_r) \quad (4.57)$$

$$D = P(D_i^{RT} = d_i^{RT}) \quad (4.58)$$

$$M^{RT} = P(m_{i,i-1}^{RT} = y_i^- | D^c = d_i^{RT} + y_i^+) P(m_{i,i+1}^{RT} = y_i^+ | D^c = d_i^{RT} + y_i^-) \quad (4.59)$$

$$p_{i^*}^j = \begin{cases} 0, & \text{si } j < i^*; \\ 1, & \text{si } j > i^* \text{ ou } i^\# \in [1, r]; \\ 1 - \frac{\bar{a}_j}{a_j}, & \text{si } j = i^*. \end{cases} \quad (4.60)$$

où pour $i = 1, \dots, r$:

$$\bar{a}_j = \max\left(a_i | L^{RT} - \sum_{l=1}^r L_l^{RT} (n_l^{RT} - d_l^{RT}) \quad (4.61)$$

$$- \sum_{l=2}^r y_l^- (L_{l-1}^{RT} - L_l^{RT}) \quad (4.62)$$

$$+ \sum_{l=1}^{r-1} y_l^+ (L_{l+1}^{RT} - L_l^{RT}) \quad (4.63)$$

$$- \sum_{l=1}^{i^*-1} a_l L_l^{RT} \quad (4.64)$$

$$- a_i L_i^{RT} > 0) \quad (4.65)$$

4.4 Résultats et analyses numériques

Nous considérons un cellule IEEE802.16e composé de deux secteurs : $r = 2$. Chaque secteur se caractérise par sa modulation de transmission : 64-QAM 3/4 ($E_2=3$ bits/symbole)

et QPSK 1/2 ($E_1=1$ bits/symbole). Le taux d'erreur block est fixé à $BLER_i = 0$. La cellule utilise 1024 sous-porteuses et les ressources totales de la cellule représentent $L = 5$. Les trafics temps réel dans le premier secteur utilisent $L_1^{RT} = 1$, et ceux du secteur extérieur utilisent $L_2^{RT} = 3$. Notons enfin que nous utiliserons ces autres paramètres : $K = 48$, $B = 2666$ symboles/seconde, $E(Pay) = 500000$ bits. Par ailleurs, le débit requis par l'ensemble des trafics temps réel correspond à $R^{RT} = 384$ kbps pour un taux de service $\mu_{RT} = 1/120$. Ces paramètres ont précédemment été étudiés et justifiés dans (Tarhini et Chahed, 2007b). Notons enfin que pour notre mécanisme de partitionnement de code : $N_{RT} = 2$. Enfin, le temps *slot* du système correspond à $T = 1$ ms. Nous précisons que les calculs relatifs à l'expression (4.37) nous ont contraint à restreindre le nombre maximum d'appels NRT dans le système à 10 appels par secteur.

4.4.1 Performances sans mobilité des utilisateurs

Cette première partie traite des résultats de cette étude vue sous un point de vue réducteur. En effet, initialement nous avons développé un modèle de performance excluant la mobilité des utilisateurs. Nous présentons ici les conclusions de cette première étude. Notons alors que pour les résultats et analyses présentés ci-dessous correspondent à une probabilité $p_{i,j}^k = 0$. Par ailleurs, cette première approche ne demandant aucune gestion des migrations d'appel, la réservation de ressources associée à cette gestion est nulle : $L_{mob} = 0$.

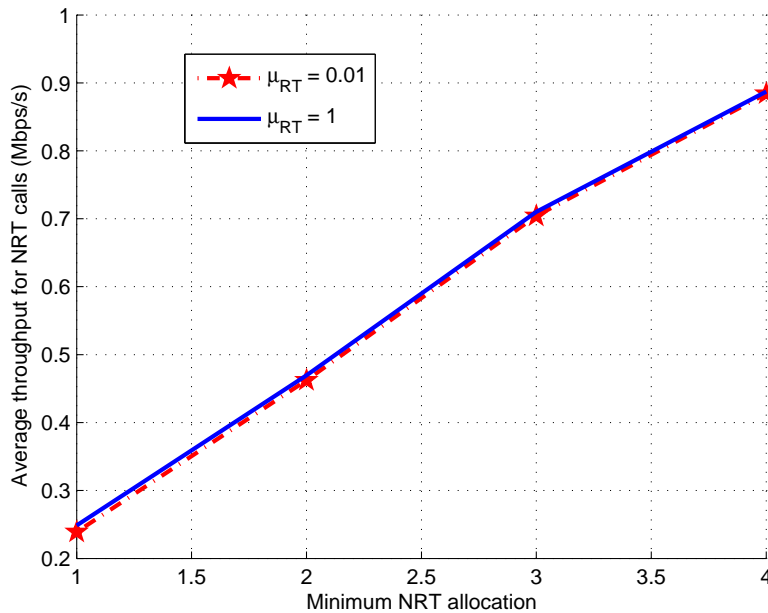


FIGURE 4.1 – Débit moyen NRT vs. bande passante NRT réservée pour différentes durées des appels RT.

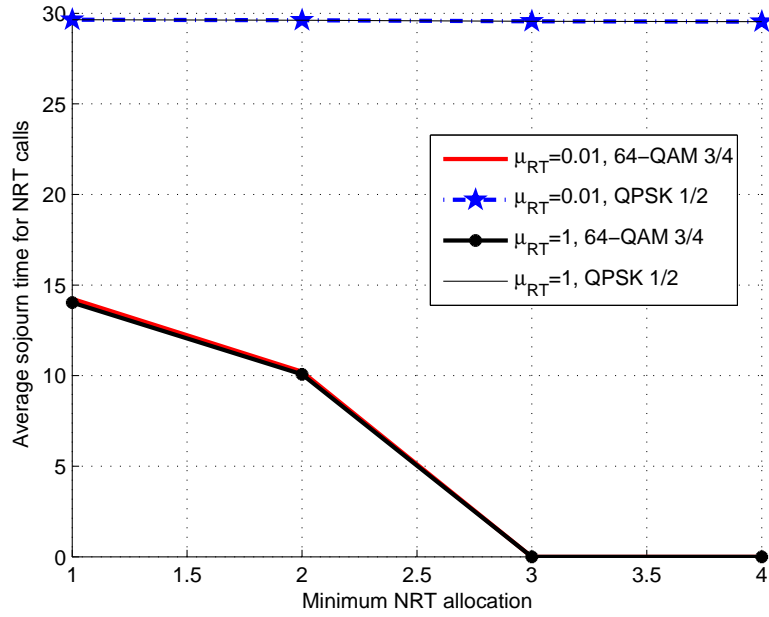


FIGURE 4.2 – Temps moyen de séjour des appels NRT vs. bande passante NRT réservée pour différentes durées des appels RT.

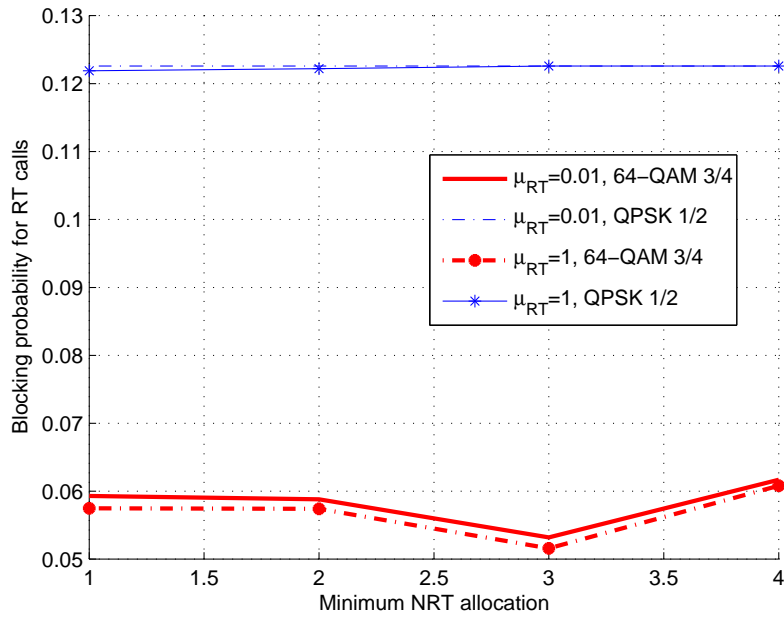


FIGURE 4.3 – Probabilité de blocage vs. bande passante NRT réservée pour différentes durées des appels RT.

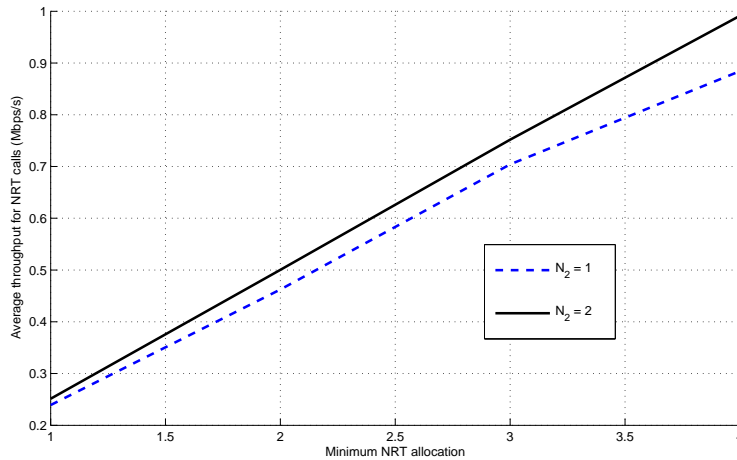


FIGURE 4.4 – Débit moyen NRT vs. bande passante NRT réservée pour différents profils de partitionnement des codes

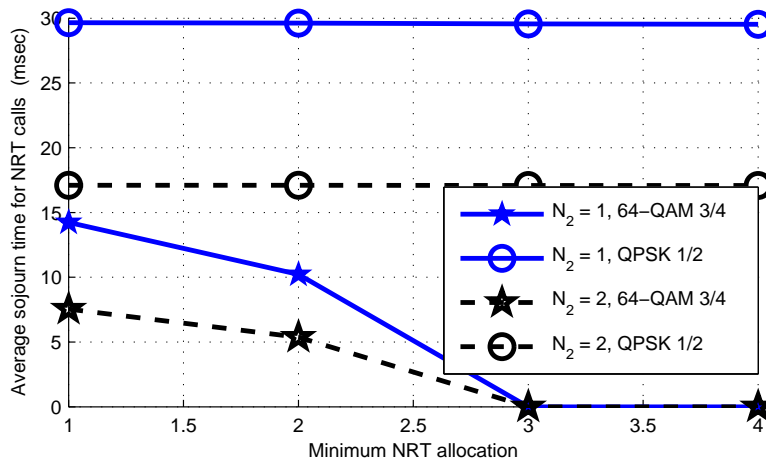


FIGURE 4.5 – Temps moyen de séjour NRT vs. bande passante NRT réservée pour différents profils de partitionnement des codes

Impact de la durée des appels temps réel

Nous visons à analyser l'influence de la durée des appels RT sur les performances du système. La figure 4.1 représente le débit moyen des appels non temps réel en fonction de la bande passante réservée à ce type de connexion. Les résultats correspondent à deux durées d'appel temps réel : $1/\mu_{RT} = 1$ et $1/\mu_{RT} = 100$.

Nous constatons dans cette figure que le débit moyen est le même dans les deux cas. Néanmoins, nous estimons que cette observation n'est valable que dans le cas où le système est constamment saturé en appel temps réel. Dès lors que l'intensité des arrivées d'appel temps réel devient faible, le débit globale des autres appels sera fonction

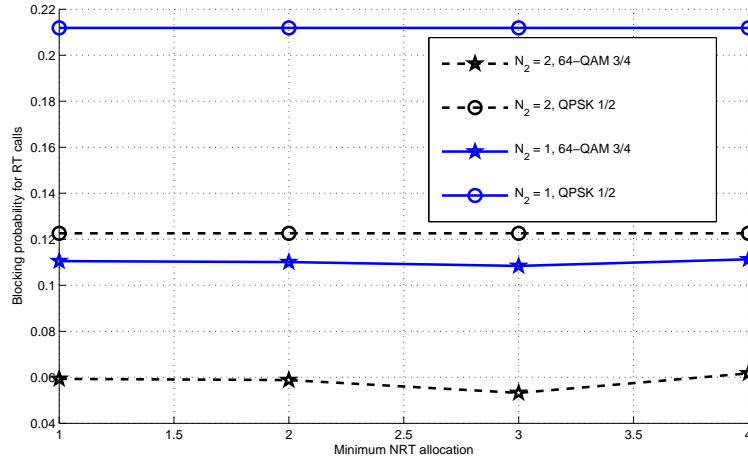


FIGURE 4.6 – Probabilité de blocage vs. bande passante NRT réservée pour différents profils de partitionnement des codes

du temps de service des appels RT. En effet, un faible temps de service diminuerait l'occupation moyenne du système en appel temps réel. De ce fait, les ressources partagées par les appels non temps réel augmenteraient. De la même manière, nous formulons ces mêmes observations sur le temps moyen de séjour : figure 4.2, et pour la probabilité de blocage : figure 4.3.

Impact du partitionnement des codes

La figure 4.4 présente le débit moyen des appels non temps réel en fonction de la bande passante qui leur sont alloués, pour différents profils de partitionnement des codes CDMA. Premièrement, nous observons que ce débit augmente linéairement par rapport à l'accroissement de la bande passante minimale allouée (L_{min}^2). Deuxièmement, l'utilisation du partitionnement des codes entraîne par contre une diminution de près de 9% du débit des appels non temps réel. En fait, en diminuant les codes disponibles aux requêtes non temps réel, nous réduisons les collisions sur les codes de requêtes d'appels temps réel, et par conséquent, nous augmentons l'occupation du système par les appels temps réel.

La figure 4.5 montre le temps moyen de transfert de données des appels non temps réel sur les deux secteurs pour différents profils de partitionnement. Comme nous pouvions le prévoir, dès lors que la bande passante minimale dédiée à ce type d'appel augmente, le temps moyen de séjour diminue fortement. De plus, nous observons sur cette figure qu'il est possible d'établir un seuil de réservation de ressource au delà duquel le gain de temps de séjour pourra être considéré comme négligeable. Par ailleurs, nous observons que l'utilisation du partitionnement des codes mène à l'effondrement des performances de temps de séjour. Comme expliqué précédemment, en réduisant le nombre de codes disponibles pour les requêtes non temps réel, nous diminuons le

débit moyen de ces appels. Du fait, le temps moyen de séjour augmente largement sur tous les secteurs de la cellule IEEE802.16e.

Finalement, la figure 4.6 représente la probabilité de blocage sur les secteurs de la cellule pour différents profils de partitionnement. En premier lieu, nous observons que la probabilité de blocage pour le secteur intérieur est sensible au partitionnement des codes entre les classes de service. Par contre, nous voyons aussi que la probabilité de blocage du secteur extérieur demeure relativement élevée et augmente doucement à mesure que cette bande passante réservée augmente. Ceci est la conséquence de l'algorithme de contrôle d'admission dans lequel le système accepte en première priorité les appels arrivant dans le secteur intérieur (c'est à dire ceux les moins gourmand en ressources). Cette observation a mené à considérer un contrôle d'admission plus flexible décrit en conclusion de ce chapitre. Enfin, la figure 4.6 montre aussi le fort impact négatif du partitionnement des codes CDMA sur la probabilité de blocage. En effet, du moment que les requêtes temps réel disposent de plus de codes dédiés, les collisions sur ces requêtes diminueront et donc un plus grand nombre d'appels temps réel se verront refusés l'accès aux ressources toujours limitées du système.

En se basant sur ces résultats, un fournisseur de service IEEE802.16e est capable de déterminer son propre seuil de bande passante allouée aux trafics non temps réel en fonction de sa politique de qualité de performance. Notons enfin que ce type de résultat est primordial pour le service de transfert de fichier de type nrtPS (voir section 1.3.5).

4.4.2 Performance en environnement mobile

Impact de la mobilité des utilisateurs

La figure 4.7, représente l'occupation moyenne du système par les appels RT et NRT. Les résultats sont donnés en fonction de la portion de bande passante L_{mob} , exprimées en pourcentage de la bande passante totale du système. La figure montre les résultats obtenus pour différents comportements de mobilité (c'est à dire pour différentes probabilités $p_{i,j}$). La figure témoigne alors que les appels NRT tirent un grand bénéfice de la bande passante conservée pour les migrations. En effet, la portion de bande passante L_{mob} permet à la fois au système de conserver des ressources en cas de migrations d'appel vers des régions extérieures. Mais elle assure surtout aux trafics NRT de pouvoir se partager une quantité appréciable de ressources implicitement sous-exploitées. Par ailleurs, l'analyse de ces résultats montre qu'un accroissement de la mobilité des utilisateurs permet d'atteindre des débits globaux plus élevés. En l'occurrence cette observation confirme les conclusions de la référence (Grossglauser, 2002) qui détermine de quelle manière la mobilité des utilisateurs accroît la capacité d'un réseau sans-fil. Au final, nous remarquons que la portion L_{mob} produit les mêmes effets que la bande passante L_{min}^{NRT} . Cette dernière vise à fournir des ressources minimum à partager entre les appels NRT. Ainsi, L_{mob} permet à la fois d'améliorer la gestion des migrations, mais aussi de répondre aux problèmes éventuels de famine de ressources pour les appels NRT.

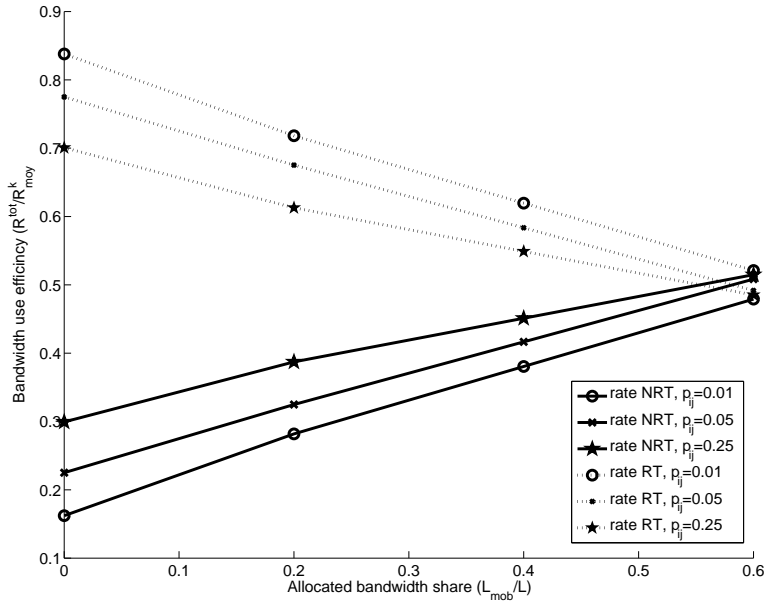


FIGURE 4.7 – Taux d'occupation des ressources des appels RT et NRT, en fonction des ressources réservées pour la mobilité et pour différents comportements de mobilité. $N_{NRT} = 2$

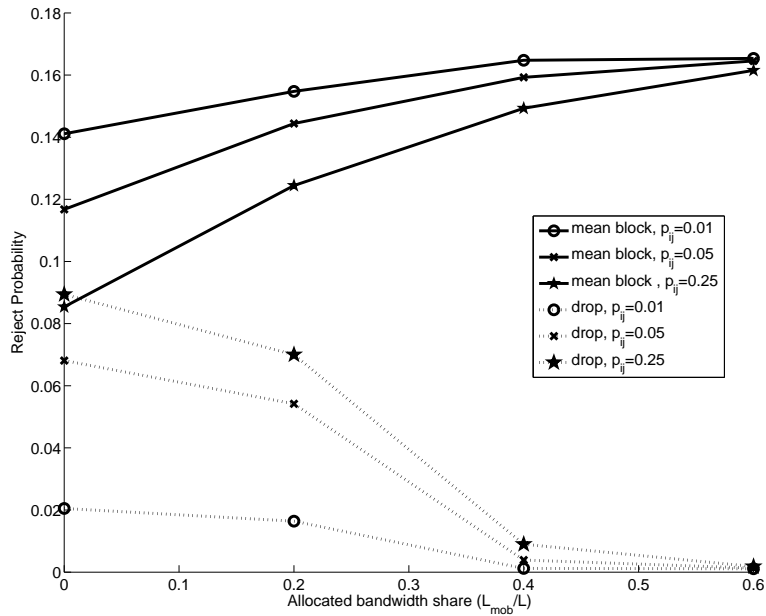


FIGURE 4.8 – Comparaison des probabilités de blocage et de perte des appels RT, en fonction des ressources réservées pour la mobilité et pour différents comportements de mobilité. $N_{NRT} = 2$

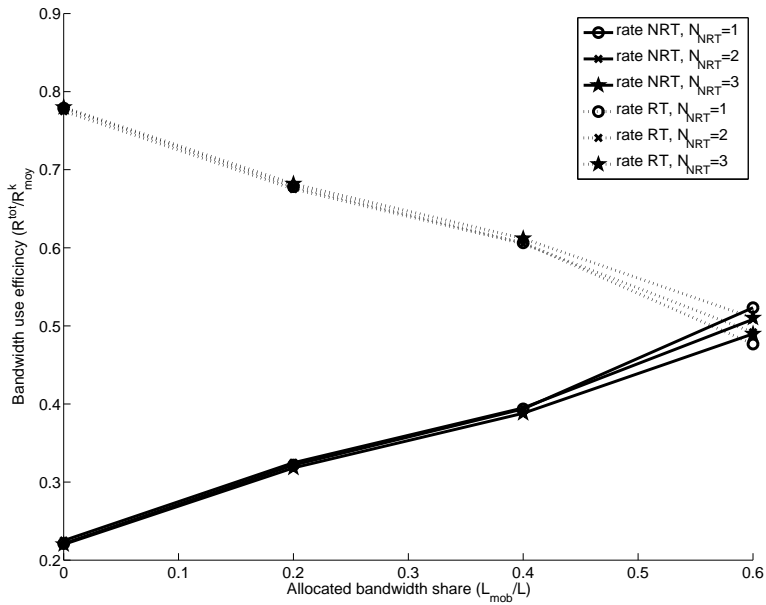


FIGURE 4.9 – Taux d’occupation des ressources des appels RT et NRT, en fonction des ressources réservées pour la mobilité et pour différents partitionnements des codes. $p_{i,j} = 0.15$

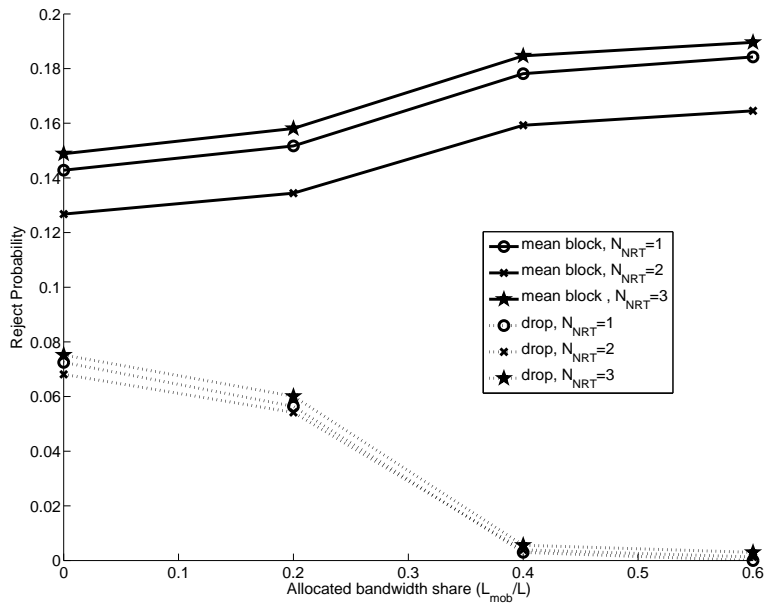


FIGURE 4.10 – Comparaison des probabilités de blocage et de perte des appels RT, en fonction des ressources réservées pour la mobilité et pour différents partitionnements des codes. $p_{i,j} = 0.15$

La figure 4.8 propose la comparaison entre la probabilité de blocage et de perte des appels RT. Ces résultats sont représentés en fonction de la bande passante réservée à la gestion des migrations ainsi que pour différents comportement de mobilité. Ainsi que nous l’attendions, la réservation de bande passante L_{mob} facilite la prise en charge des migrations d’appel vers les secteurs extérieurs. La probabilité de perdre un appel en migration se retrouve alors fortement réduite à mesure que la partie réservée augmente. D’ailleurs, cette réservation de bande passante pour les migrations permet de conserver une probabilité de perte d’appel relativement faible alors que les utilisateurs expriment une mobilité extrêmement forte : la figure montre que la probabilité de perte se retrouve multipliée par 7 alors que les utilisateurs sont 25 fois plus rapides. De plus, nous pouvons apprécier sur cette figure l’évolution de la probabilité de blocage des appels en entrée du système. Cette probabilité atteint un plateau à mesure que la réservation de ressources L_{mob} augmente linéairement. De ce fait nous concluons ici que cette portion L_{mob} permet de réduire de manière rapide la probabilité de perte des appels tout en conservant une probabilité de blocage acceptable ou, au pire, caractérisée par un plateau de valeurs quasi-invariantes.

Impact du partitionnement des codes

La figure 4.9 représente l’occupation moyenne du système par les appels RT et NRT. Les résultats sont donnés en fonction de la portion de bande passante L_{mob} , exprimés en pourcentage de la bande passante totale du système. La figure montre les résultats obtenus pour différents schémas de partitionnement des codes alloués à chaque type d’appel (N_{NRT}). De même que précédemment, nous observons en premier lieu que le débit des appels NRT profite de la mobilité des utilisateurs. Par contre, ce débit paraît insensible aux configurations possibles du schéma de partitionnement. Ce fait s’explique dès lors que les schémas de partitionnement permettent uniquement de faciliter l’accès aux système des appels NRT. Or la portion de bande passante accessible aux appels NRT dépend uniquement du nombre des appels RT dans le système. Ainsi, la capacité des appels NRT à profiter de la totalité de cette bande passante ne dépend nullement de la quantité d’appels NRT ayant eu accès à cette ressource.

La figure 4.10 propose la comparaison entre la probabilité de blocage et de perte des appels RT. Ces résultats sont représentés en fonction de la bande passante réservée à la gestion des migrations ainsi que pour différents schémas de partitionnement des codes alloués à chaque type d’appel (N_{NRT}). Nous observons que la probabilité de blocage, au contraire de celle de perte des appels migrant, peut être réduite par une configuration efficace du partitionnement des codes entre les types de service. Par exemple, la figure témoigne d’un taux de blocage minimum pour un partitionnement caractérisé par $N_{NRT} = 2$. En fait, cette configuration correspond à la configuration optimale définie dans le chapitre 3. Nous y précisons que la répartition des codes entre types de services dépendait du nombre d’utilisateurs moyen de ces services. Dès lors, du moment que la répartition de code peut se faire de manière optimale, il nous est alors possible de prétendre à la meilleur optimisation des paramètres du système assurant une probabilité de blocage initiale la plus basse possible via l’utilisation des codes, et la

probabilité acceptable de perte d'appel la plus haute via la portion de bande passante réservée aux appels migrant.

4.5 Bilan et perspectives

Dans ce chapitre, nous définissons un modèle pour cellule IEEE802.16e. La cellule se compose de différents secteurs de modulation suivant le principe de l'AMC. Les utilisateurs suivent le comportement établi dans le chapitre précédent. De plus, nous y intégrons le principe de partitionnement en plus des mécanismes MAC définis par le standard.

Du moment qu'aucun algorithme standard de contrôle d'admission des connexions entrantes n'est fourni directement par la norme IEEE802.16e, nous désirions définir différentes approches de compromis de différenciation des deux principaux types de service. L'objectif final était d'en extraire des éléments d'amélioration conjointe de leur performances respectives. Au travers de ce dernier travail, nous proposons un algorithme de contrôle d'admission où les appels non temps réel profitent d'une quantité minimale de ressources à se partager équitablement entre eux, ainsi que d'une partie de bande passante visant à empêcher la perte des appels temps réel. Ces derniers, en migrant dans des secteurs extérieurs, demandent un surplus de bande passante pour la pérennité de leur services.

La station de base exploite un nouvel algorithme de contrôle d'admission. Il fournit aux trafics temps réel un débit fixe. Les autres types d'appel se partagent alors les ressources laissées disponibles. Durant cette étude, nous caractérisons les relations liées à la probabilité de blocage des appels temps réel, au débit moyen des appels non temps réel ainsi qu'à leur temps moyen de séjour. Au cours de l'analyse numérique, nous observons le comportement des performances du système dans un contexte de régime saturé, où l'intensité d'arrivée des appels temps réel est forte.

Dans un premier temps, nous nous intéressons au comportement du système dans le cas où les utilisateurs ne sont pas mobiles. Dans ce cas, nous observons l'influence des changements de taux de services des appels temps réel. Sur ce point, nos résultats témoignent que les différentes métriques de performances étudiées ne profitent pas des variations de ce taux de service. Par contre, nous jugeons qu'un gain significatif serait observé du moment que le système ne demeure pas continuellement saturé en appels de type temps réel. L'analyse précise de ce fait nous semble secondaire a vu des objectif premier de cette étude. Aussi, nous avons réservé cette analyse à des travaux futurs.

Ensuite, nous étudions l'impact du principe de partitionnement des codes entre les appels temps réel et non temps réel. Le débit global des appels non temps réel subit une diminution notable dès lors que les appels temps réel disposent d'une plus grande plage de codes pour engager leur requêtes de ressource. En effet, une telle augmentation de code facilite l'accès aux ressources du système pour les appels temps réel. Ainsi, ceci a pour conséquence direct d'augmenter l'occupation globale du système par ce type d'appels. Cette augmentation générale réduit donc les ressources dont disposent

les appels non temps réel. En parallèle à cela, nous observons les gains importants obtenus par l'intégration au CAC d'une quantité de ressources réservée aux appels non temps réel. Ainsi, dans la mesure où le principe de partitionnement serait largement défavorable à ces derniers, une légère augmentation de cette quantité réservée permettrait de compenser la perte de débit. En outre, nous observons naturellement ce même comportement vis à vis du temps de séjour des appels non temps réel. De plus, ces derniers résultats révèlent aussi l'existence d'un critère de réservation optimal. En effet, en fonction du nombre d'appels non temps réel en cours et de la taille moyenne des paquets qu'ils transmettent, il est possible de déterminer une quantité optimale de ressources à réserver. Au delà de cette quantité, la diminution du temps de séjour devient alors négligeable, voir nulle. De l'ensemble de ces observations, nous estimons qu'il est envisageable d'établir une logique d'adaptation dynamique de cette bande passante réservée en fonction de l'état du système. En ce qui concerne la probabilité de blocage, là aussi, le principe de partitionnement a un impact majeur. L'augmentation du nombre de codes dédiés pour les appels temps réel facilite leur accès au canal. Dans la mesure où le système subit déjà une lourde charge, ce fait aboutit à une augmentation considérable de la probabilité de collision. Cet fait confirme une des observations faites en conclusion du chapitre précédent (voir section 3.4). En effet, le principe de partitionnement offre la possibilité d'établir un premier CAC décrit comme « virtuel ». Dans un contexte de forte charge, la station de base peut diminuer la plage des codes dédiés aux appels temps réel. Ce fait, bien qu'augmentant le délai d'établissement de ces services, aboutit à une baisse appréciable de la probabilité de blocage qu'ils subissent. Aussi, nous encourageons l'étude d'un algorithme de CAC et de partitionnement entièrement dynamique où la répartition des codes et la réservation de ressources sont fonction de l'état du système et des performances de connexion et de transmission qui en découlent.

Afin de répondre à la critique établie ci-dessus, nous identifions une perspective principale d'étude : à l'algorithme étudié ici doit s'adjoindre une répartition des acceptations d'appel sur les différents secteurs de la cellule. Afin de limiter les rejets massifs des appels périphériques, nous voudrions intégrer une probabilité d'acceptation des appels entrant propre à chaque secteur. Aussi nous désirons définir un paramètre α_i , $i \in [1, \dots, r]$ définissant la probabilité qu'un appel arrivant en secteur i soit accepté dans le système du moment que ce dernier dispose encore de ressources suffisantes. Ainsi, la station de base n'accepterait dans une première étape qu'une partie des appels dans chacun des secteurs. A la suite de cela, si des ressources demeurent libres et des appels en attente, le système reprend alors une acceptation progressive des appels sur les secteurs. Ce principe par étape s'opérerait alors jusqu'à surcharge du système ou satisfaction de tous les appels entrants. Dès lors, une telle étude viserait à définir ces probabilités d'acceptation capables d'assurer un blocage acceptable et équitable pour l'ensemble des secteurs de la cellule. Par contre, il faut prendre en considération qu'une telle option ne permet plus d'assurer la gestion la plus optimale possible de la ressource radio.

Enfin, nous soulignons le fait qu'un tel algorithme de contrôle d'admission aboutit à une forte probabilité de blocage des appels dans les secteurs extérieurs. En fait, ce type

d'algorithme prend uniquement en considération l'objectif d'efficacité de la gestion de la ressource radio. En effet, en acceptant en première priorité les appels les moins gourmands, nous assurons à la station de base une exploitation optimale des sous-porteuses allouées aux appels entrant dans le système. Ainsi cet algorithme de CAC permet la prise en charge d'un maximum de communications. Néanmoins, il est important d'en extraire une critique majeure. Dans un contexte d'utilisation commerciale, les niveaux de blocages obtenus en périphérie de cellule aboutiront inévitablement à un fort mécontentement des utilisateurs. Ainsi, même si l'opérateur est à même, par cet algorithme, de satisfaire le nombre maximum de clients, il devra faire face à son incapacité à satisfaire les clients en périphérie de son infrastructure.

Dans un second temps, nous considérons les utilisateurs, et donc leur appels, comme mobiles. Dans ce contexte d'arrivée, de départ et de migration en temps discret, le système vise à accepter en premier les appels les moins gourmands en ressources, c'est à dire ceux des secteurs les plus internes à la cellule IEEE802.16e. Les appels temps réel se caractérisent par un débit constant au cours du temps et un temps de service indépendant des ressources consommées. La bande passante qui leur sont allouée est alors dépendante de la modulation du secteur duquel il sont issus et le temps de services est uniquement fonction du comportement de l'utilisateur. A l'inverse, les appels non temps réel n'ont aucune exigence de débit et supportent la variation dynamique des bandes passante qui leur sont allouées. Ils se partagent équitablement les sous-porteuses laissées libres par les appels temps réel, et leur débits respectifs dépendent de la modulation qu'ils exploitent. Enfin, leur temps de service dépendent directement des ressources qui leur sont attribuées au cours du temps.

Les résultats montrent que l'existence d'une bande passante facilitant la migration des appels temps réel vers les secteurs extérieurs réduit grandement la probabilité de perte de ces appels sans pour autant augmenter de manière significative la probabilité de bloquer les nouvelles arrivées de ces appels. Par ailleurs, le fait que les appels non temps réel puissent disposer de tout ou partie de cet espace de sous-porteuses leur assure un approvisionnement minimal en ressources de transmission. Aussi, nous concluons en préconisant l'utilisation seule de cette espace L_{mob} . Quel que soit l'importance de la mobilité des utilisateurs, cet espace de ressource est à même de répondre aux objectifs ayant initialement motivés la mise en place de l'espace L_{min}^{NRT} . Un dernier élément de conclusion confirme qu'une mobilité intense des utilisateurs permet d'améliorer la capacité globale du système. Nous précisons d'ailleurs que ceci est vrai dans le « WiMAX » dès lors que les trafics de signalisation induis par la migration circule sur des canaux physique séparés de l'espace OFDMA (temps-fréquence) servant à la transmission des données. Finalement, nos éléments de conclusions précisent les critères à prendre en considération quant à l'optimisation des grandeurs étudiées dans ce chapitre ainsi que des perspectives et moyens de priorisation des types de service entre eux.

Nos travaux futurs sont maintenant motivés par une observation majeure : la présente étude nous invite naturellement à étendre le modèle en intégrant une mobilité inter-cellulaire. Le standard IEEE802.16e assure maintenant la mobilité des utilisateurs sur plusieurs cellules en se basant sur le principe de *handover*. Aussi, nous engageons

maintenant leur intégration dans le modèle. D'ailleurs, les premiers résultats obtenus à ce jour nous montrent que le modèle défini dans ce chapitre permet une intégration extrêmement aisée des événements propres aux arrivées et départs extra-cellulaires intervenants dans un environnement à temps discret.