Chapitre 11

# Application de technique d'apprentissage dans les réseaux mobiles

## 11.1. Introduction

Avec l'évolution de la société actuelle, les personnes se déplacent de plus en plus, tout en ayant besoin de communiquer pendant leurs déplacements. Ce phénomène a provoqué une demande accrue et orienté les études vers le développement de systèmes très sophistiqués afin de répondre aux nouveaux besoins des utilisateurs. Ces besoins ont effectivement changé ; si la voix était à l'origine le seul besoin, la demande en transmissions sans fils fournissant des communications fiables de son à haute définition, d'images, voire de vidéos de haute qualité est devenue de plus en plus prisée par un nombre croissant d'utilisateurs. Ces derniers, souhaitent que la mobilité soit complètement transparente afin de bénéficier de performances comparables à celles des réseaux filaires, en dépit de la gourmandise en bande passante de ces nouveaux services.

Les systèmes cellulaires sont, sans aucun doute, ceux qui ont connu la plus grande évolution ces dernières années. La zone géographique desservie par un réseau cellulaire est divisée en petites surfaces appelées cellules. Chacune d'elles est couverte par un émetteur appelé « station de base » (*Base Station* - BS). La bande passante dans un tel réseau est divisée en un ensemble disjoint de canaux radios dont le type dépend de la technique d'accès utilisée. Ces canaux peuvent être utilisés simultanément, à condition de maintenir une qualité du signal radio acceptable. Cette division peut se faire grâce à différentes techniques d'accès telles que FDMA (*Frequency-Division Multiple Access*), TDMA (*Time-Division Multiple Access*),

---

Chapitre rédigé par SIDI-MOHAMMED SENOUCI.

CDMA (*Code-Division Multiple Access*) ou toutes combinaisons de ces méthodes [CAL 92], [GIB 96], [AGH01]. Il reste maintenant à définir la manière dont sont attribués ces canaux aux cellules. Il existe principalement deux méthodes : FCA (*Fixed Channel Allocation*), où chaque cellule possède un nombre fixe de canaux et DCA (*Dynamic Channel Assignment*) où tous les canaux sont groupées dans un pool (ou groupe) commun et sont assignés aux cellules de manière dynamique [KAT 96]. Dans les systèmes de type FCA ou DCA, un canal libre ne violant pas la contrainte de réutilisation de canaux[1] est alloué à chaque usager. Mais quand ce dernier passe d'une cellule à une autre, il doit demander un nouveau canal libre dans la cellule destination. Cet événement, appelé transfert inter-cellulaire ou « *handoff* », doit être transparent à l'utilisateur. Si la cellule destination n'a aucun canal disponible, l'appel est coupé. Un des grands challenges pour ce type de réseaux est la gestion de cette mobilité des usagers durant une même communication. En effet, la disponibilité des ressources radios durant toute la durée de la communication n'est pas nécessairement garantie, et ces utilisateurs peuvent subir une dégradation ou même une rupture de la communication lors d'un transfert inter-cellulaire.

L'un des soucis majeurs lors de la conception des réseaux cellulaires est la réduction de la probabilité de coupure. En effet, du point de vue de l'usager, elle est beaucoup plus désagréable qu'un échec de connexion. Ceci est d'autant plus important qu'afin de répondre à la croissance des réseaux cellulaires, la taille des cellules est de plus en plus réduite, ce qui augmente considérablement le nombre de transferts inter-cellulaires. Ainsi, étant donné que la ressource radio est une ressource rare, il est impératif de l'utiliser au maximum et particulièrement dans le cas d'un réseau cellulaire multiservices supportant plusieurs classes de trafic dont chacune demande un niveau de QoS différent. Pour un opérateur il est parfois préférable de bloquer un appel d'une classe de service moins prioritaire (données par exemple) et d'accepter un autre appel d'une classe plus prioritaire (voix par exemple). Par conséquent, une bonne politique de contrôle d'admission d'appel (CAC) est certainement nécessaire afin de permettre de maximiser l'utilité de l'ensemble de ces ressources radios. Pour accomplir cet objectif, il est également indispensable de trouver une bonne méthode d'allocation de la totalité de la bande passante disponible à l'ensemble des cellules. Ces nouveaux mécanismes (CAC, allocation dynamique des ressources) doivent également faire face aux changements fréquents des conditions de trafic dans les réseaux cellulaires.

L'objectif de ce chapitre est de prouver qu'il est possible d'utiliser des techniques venant du monde de l'Intelligence Artificielle (IA), et plus spécialement des techniques d'apprentissage afin d'élaborer des mécanismes robustes et très efficaces pour résoudre les problématiques rencontrées dans ces réseaux cellulaires.

---

1. Un canal peut être utilisé dans plusieurs cellules tant que la contrainte d'interférences est respectée.

Ces mécanismes doivent également exploiter l'expérience et la connaissance qui pourraient être acquises en cours de fonctionnement du réseau.

Pour ceci, nous avons développé de nouveaux mécanismes de contrôle d'admission d'appels dans un réseau cellulaire en considérant les deux schémas d'allocation de canaux : fixe (FCA) et dynamique (DCA). Nous avons également développé un nouveau mécanisme d'allocation dynamique de ressources permettant de choisir le meilleur canal parmi tous les canaux disponibles dans le pool commun, dans un objectif de maximiser le taux d'utilisation de tous les canaux. Ces solutions sont obtenues en utilisant l'algorithme d'apprentissage par renforcement Q-learning [WAT 89], [WAT 92].

Le reste du chapitre est organisé comme suit. La section 2 présente brièvement la notion d'apprentissage, en mettant l'accent sur l'apprentissage par renforcement et son application dans le domaine des réseaux de télécommunication. La section 3 présente une nouvelle méthode de contrôle d'admission des appels dans les réseaux cellulaires basée sur l'apprentissage par renforcement. La section 4 expose une nouvelle politique d'allocation dynamique des ressources radios dans les systèmes cellulaires, basée également sur l'apprentissage par renforcement. Finalement, la section 5 conclut le chapitre.

## 11.2. L'apprentissage

Le terme apprentissage désigne la capacité à organiser, à construire et à généraliser des connaissances pour une utilisation ultérieure. C'est donc la capacité à tirer profit de l'expérience pour améliorer la résolution d'un problème. Selon le type d'informations disponibles, deux grandes catégories d'approches peuvent être distinguées. La première qualifiée d'apprentissage non-supervisé vise à regrouper des objets en classes, en se basant sur des ressemblances entre eux. La deuxième approche est l'apprentissage supervisé, basée quant à elle sur un ensemble d'apprentissage constitué d'objets dont la classe est connue à priori.

### 11.2.1. *Apprentissage non-supervisé*

L'apprentissage non-supervisé, appelé aussi apprentissage à partir d'observations, consiste à déterminer une classification à partir d'un ensemble d'objets ou de situations données. On dispose d'une masse de données indifférenciées, et l'on désire savoir si elles possèdent une quelconque structure de groupes. Il s'agit d'identifier une éventuelle tendance des données à être regroupées en classes. Ce type d'apprentissage appelé '*Clustering'*, est retrouvé dans la classification automatique et dans la taxinomie numérique. Il recherche des

régularités parmi un ensemble d'exemples, sans être nécessairement guidé par l'utilisation qui sera faite des connaissances apprises. Il regroupe l'ensemble des exemples de manière à ce que les exemples au sein d'un même groupe se ressemblent suffisamment, et que les exemples de groupes différents soient suffisamment différents.

### 11.2.2. *Apprentissage supervisé*

Dans ce type d'apprentissage, un maître (ou superviseur, d'où le nom d'apprentissage supervisé) fournit soit l'action qui devrait être exécutée, soit un gradient sur l'erreur commise. Dans les deux cas, le maître fournit au contrôleur une indication sur l'action qu'il devrait générer afin d'améliorer ses performances. L'utilisation d'une telle approche présuppose l'existence d'un expert capable de fournir un ensemble d'exemples, appelé base d'apprentissage, formés de situations et d'actions correctes associées. Ces exemples doivent être représentatifs de la tâche à accomplir.

L'une des variantes de l'apprentissage supervisé, dans lequel une « critique » de la réponse calculée est fournie au réseau, est l'apprentissage par renforcement (*Reinforcement Learning* – RL). C'est cette variante d'algorithmes, détaillée ci-après, qui nous a paru la plus adaptée pour résoudre les quelques problèmes liés aux réseaux cellulaires et traités dans ce chapitre.

### 11.2.3. *Apprentissage par renforcement*

L'apprentissage par renforcement (dit aussi apprentissage semi-supervisé) est une variante de l'apprentissage supervisé [SUT 98]. Par opposition à l'approche supervisée, l'agent maître dans l'apprentissage par renforcement a un rôle d'évaluateur et non pas d'instructeur. Il est en général appelé critique. Le rôle du critique est de fournir une mesure indiquant si l'action générée est appropriée ou non. Il s'agit de programmer un agent au moyen d'une évaluation par pénalité/récompense sans avoir besoin de spécifier comment la tâche doit être remplie. Dans ce cadre, on doit indiquer au système quel est le but à atteindre, et celui-ci doit apprendre par une succession d'essais/erreurs (en interaction avec l'environnement) comment atteindre le but fixé.

Les composantes de l'apprentissage par renforcement sont l' « apprenti » agent, l'environnement où il agit ainsi que la tâche qu'il doit réaliser (cf. Figure 11.1). L'interaction entre l'agent et l'environnement est continue. D'une part le processus de décision de l'agent choisit des actions selon les situations perçues de l'environnement, et d'autre part ces situations évoluent sous l'influence de ces

actions. Chaque fois que l'agent effectue une action, il reçoit une récompense. Celle-ci est une valeur scalaire indiquant à l'agent la conséquence de cette action.

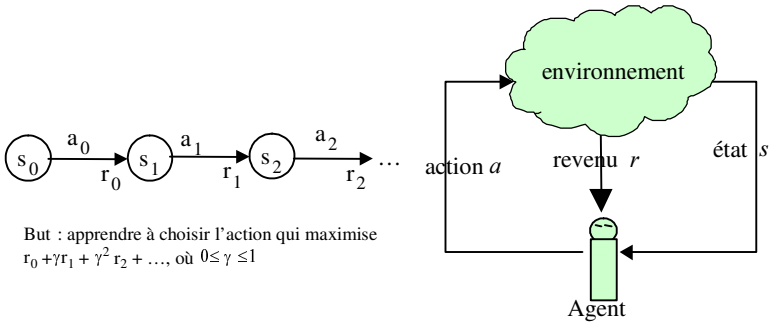


**Figure 11.1.** *Interaction Agent-Environnement.*

Pour être plus formel, dénotons $s$ ($\in S$, un ensemble fini), une représentation de l'état actuel de l'environnement, $a$ ($\in A$, un ensemble fini) l'action choisie et $r$ ($\in R$, un ensemble fini) la récompense reçue. L'interaction entre l'agent et l'environnement consiste, à chaque instant, en les séquences suivantes :

– l'agent observe l'état actuel de l'environnement $s_t \in S$ ;

– en se basant sur l'état $s_t$, l'agent prend une décision en exécutant une action $a_t \in A$ ;

– l'environnement fait alors une transition vers un nouvel état $s_{t+1}=s' \in A$ suivant la probabilité $P_{ss'}(a)$ ;

– l'agent reçoit instantanément un certain revenu $r_t=r(s_t,a_t)$ indiquant la conséquence de cette décision.

Le processus de décision de l'agent s'appelle politique et c'est une fonction de l'ensemble des états vers l'ensemble des actions ($\pi : S \rightarrow A$). L'agent doit apprendre une politique, $\pi$, permettant de choisir la prochaine action $a_t=\pi(s_t)$ à effectuer, et ceci en fonction de l'état actuel $s_t$. L'interaction entre l'agent et l'environnement est continue et l'apprenti agent modifie sa politique selon son expérience et selon le but consistant à maximiser le cumul des récompenses dans le temps. Ce cumul $V_\pi(s_t)$, réalisée en suivant une politique arbitraire $\pi$, à partir d'un état initial $s_t$, est définie comme suit :

$$V^{\pi}(s) = E\left\{ \sum_{t=0}^{\infty} \gamma^t r(s_t, \pi(s_t)) \middle| s_0 = s \right\} \qquad [11.1]$$

où $E$ désigne l'opérateur espérance et le facteur $0 \leq \gamma \leq 1$ représente la constante de propagation temporelle.

L'objectif, pour l'agent, est donc de maximiser cette somme des renforcements reçus, et son apprentissage s'effectue par de nombreuses expériences. L'agent est guidé en cela par divers algorithmes présentés ci-après.

11.2.3.1. *Méthodes de résolution*

Il existe trois classes fondamentales de méthodes permettant de résoudre un problème d'apprentissage par renforcement : la programmation dynamique PD, les méthodes de Monte Carlo MC, et l'apprentissage par différences temporelles TD [COR]. Chaque classe possède des avantages et des inconvénients. La programmation dynamique possède des fondements mathématiques bien connus/étudiés mais nécessite un modèle complet et précis de l'environnement. Les méthodes MC ne nécessitent pas de modèle et sont conceptuellement simples, mais sont inadaptées à un calcul pas à pas incrémentiel. Enfin, l'approche TD combine les deux premières méthodes, et récupère ainsi la meilleure part de chacune. Cette approche ne nécessite pas de modèle et est incrémentale. Ces méthodes se distinguent aussi en matière d'efficacité et de rapidité de convergence. Nous décrivons, dans ce qui suit, deux méthodes de résolution d'apprentissage par différences temporelles : le Q-learning et Sarsa.

11.2.3.1.1. L'algorithme Q-Learning

Le Q-learning, développé en 1989 par Watkins [WAT 92], [WAT 89], fait partie des méthodes d'apprentissage par renforcement sans modèle puisqu'il s'agit d'apprendre par l'expérience les actions à effectuer en fonction de l'état actuel. L'objectif de l'agent est donc d'apprendre une politique $\pi : S \rightarrow A$, permettant de choisir la prochaine action $a_t=\pi(s_t)$ à effectuer et ceci en fonction de l'état actuel $s_t$.

Pour chaque politique $\pi$, on associe une valeur $Q^{\pi}(s,a)$, qu'on appellera sa Q-valeur et qui représente la moyenne des gains prévue si l'action $a$ est exécutée quand l'état courant du système est $s$ et qu'ensuite $\pi$ est adoptée comme politique de décision. La politique optimale $\pi^*(s)$ est la politique qui maximise le cumul des revenus $r_t=r(s_t,a_t)$ reçus après un temps infini. Le but de l'algorithme du Q-learning est de rechercher une approximation pour $Q^*(s,a)=Q^{\pi^*}(s,a)$ de manière récursive avec seulement le quadruplet $(s_t, a_t, s'_t, r_t)$ comme information disponible. Cette information comprend l'état à l'instant $t$ ($s_t$), l'état à l'instant $t+1$ ($s'_t=s_{t+1}$), l'action prise quand le système est à l'état $s_t$ ($a_t$) et le revenu reçu à l'instant $t$ ($r_t$) suite à l'exécution de cette action.

Les Q-valeurs sont mises à jours, de façon récursive à chaque transition, en utilisant la formule [11.2] suivante :

$$Q_{t+1}(s,a) = \begin{cases} Q_t(s,a) + \alpha_t \Delta Q_t(s,a), & si\ s = s_t\ et\ a = a_t \\ Q_t(s,a), & sinon \end{cases} \quad [11.2]$$

où

$$\Delta Q_t(s,a) = \left\{ r_t + \gamma \max_b \left[ Q_t(s'_t, b) \right] \right\} - Q_t(s,a) \quad [11.3]$$

D'où l'algorithme récapitulé dans la Figure 11.2 suivante.

```
Initialiser Q0(s, a) à des valeurs aléatoires
Choisir un point de départ s0
tant que la politique n'est pas suffisamment bonne
choisir at en fonction des valeurs Qt(st,.)
at = f(Qt(st,.))
obtenir en retour : st+1 (s') et rt
mettre à jour Qt+1(st, at) en utilisant la formule [11.2]
Fin tant que
```

**Figure 11.2.** *Algorithme Q-learning.*

Il est alors nécessaire de régler le coefficient $\alpha$ afin de fixer progressivement la politique apprise. Le facteur $\gamma$ permet, pour sa part, de moduler l'importance des récompenses escomptées à venir. Dans [WAT 92], les auteurs démontrent que si chaque paire $(s,a)$ est infiniment visitée, et que le taux d'apprentissage $\alpha$ tend vers zéro, $Q_t(s,a)$ converge, lorsque $t \to \infty$, vers $Q^*(s,a)$ avec une probabilité de 1. La politique optimale sera, alors, celle avec la plus grande Q-valeur :

$$\pi^*(s) = \arg \max_{a \in A(s)} Q^*(s,a)$$

Le choix de l'action (la fonction *f*) n'est pas décrit. Il est possible d'imaginer différentes stratégies de sélection (exploration), par exemple l'action qui a été la moins utilisée, ou celle qui renvoie la plus forte Q-valeur. Dans nos travaux, nous avons testé un ensemble de méthodes d'exploration, telles que $\in$-*gloutonne* (ou $\in$-

*directed*), *Boltzmann* et même des stratégies aléatoires que nous détaillerons ultérieurement.

11.2.3.1.2. L'algorithme Sarsa

Sarsa [SUT 98] est un autre algorithme de résolution permettant de résoudre un problème d'apprentissage par renforcement. Contrairement au Q-learning, lors de la mise à jour des Q-valeurs, la politique utilisée pour le choix de l'action $a$ à l'instant $t$ est la même que celle utilisée pour le choix de l'action $b$ à l'instant $t+1$, i.e. :

$$\Delta Q_t(s,a) = \left\{ r_t + \gamma\, Q_t(s'_t,b) \right\} - Q_t(s,a)$$

La politique d'exploration utilisée peut être $\in$-gloutonne par exemple. Si chacune des paires $(s,a)$ est infiniment visitée, Sarsa converge vers la politique optimale.

11.2.3.2. *Applications de techniques d'apprentissage par renforcement – état de l'art*

Nous allons énumérer, ci-dessous, quelques travaux relatifs à des applications classiques (robotique, jeux, etc.) et à quelques applications réseaux (routage, CAC, etc.) utilisant l'apprentissage par renforcement comme solution. En effet, les travaux relatifs à des applications réseaux de télécommunications ne sont malheureusement pas aussi nombreux.

Les premiers travaux ont été effectués dans le cadre de l'apprentissage de jeu à deux joueurs (jeu de dames, jeu du Backgammon, etc.). L'autre grand domaine d'applications de l'apprentissage par renforcement est celui de la robotique autonome, un domaine très peu abordé avec l'apprentissage supervisé du fait de l'impossibilité de modéliser le monde réel avec suffisamment de précision pour tenir compte de l'hétérogénéité des capteurs, du bruit ambiant et la dynamique robot-monde extérieur. Dans [SRI 00], les auteurs proposent un système multi-agents (SMA) capable de prendre des décisions économiques telles que fixer les prix dans un contexte de marché compétitif. Ils démontrent qu'avec le Q-learning, le système permet de trouver la politique de prix optimale et de diminuer le phénomène de 'guerre des prix' entre les différents vendeurs.

Les travaux [BOY 94], [MAR 00], [MAR 98] sont des propositions de routage dans les réseaux de télécommunications, en utilisant des techniques d'apprentissage par renforcement comme solution. Les auteurs dans [BOY 94] proposent une extension de l'algorithme à vecteur de distance de Bellman-Ford (DBF) et qu'ils ont appelé Q-routing. Le module d'apprentissage par renforcement est intégré dans chaque nœud d'un réseau de commutation. La politique de routage tente de trouver

le meilleur nœud adjacent permettant d'atteindre la destination avec un « temps de transmission » minimum.

Nous avons proposé, dans [SEN3 03], un protocole de routage ad hoc Q-AOMDV basé sur l'un des plus importants protocoles de routage ad hoc actuels qui est AODV (Ad hoc On Demand Distance Vector) [MANET]. L'objectif dans Q-AOMDV est d'équilibrer la consommation de l'énergie à travers tout le réseau, en envoyant le trafic de données sur des routes différentes. Le choix de la meilleure route est réalisé en utilisant une adaptation de l'algorithme Q-routing. Rappelons que AODV est un protocole de routage ad hoc à la demande, qui ne maintient qu'une seule route vers la destination. Contrairement à AODV standard, l'ensemble des résultats montre que Q-AOMDV équilibre la consommation de l'énergie sur la totalité du réseau ad hoc tout en évitant de partitionner le réseau en sous-réseaux disjoints.

Une proposition particulièrement intéressante traitant de l'allocation de ressources est celle proposée par *Nie* et *Haykin* [NIE 99]. Nous nous intéressons plus spécialement à cette proposition puisqu'elle fait l'objet d'une extension abordée dans la section 4. Les auteurs proposent de résoudre le problème d'allocation dynamique de ressources dans un réseau cellulaire de type GSM (service voix uniquement) en utilisant le Q-learning.

Enfin, les problèmes de contrôle d'admission CAC dans les réseaux fixes et utilisant l'apprentissage par renforcement ont été traités dans quelques travaux [MAR 00], [MAR 98], [MAR 97], [RAM 96], [MIT 98]. Les auteurs de [MAR 00] proposent, par exemple, de résoudre ce problème dans un réseau à intégration de services tel que ATM (Asynchronous Transfer Mode). Les auteurs se mettent dans le contexte où un fournisseur de services désire vendre ses ressources réseaux dans le but de maximiser l'ensemble de ses revenus.

## 11.3. Contrôle d'admission des appels

Cette section présente une approche permettant de résoudre le problème du contrôle d'admission des appels dans les réseaux cellulaires de type FCA supportant plusieurs classes de trafic.

Nous rappelons qu'une zone géographique desservie par un réseau mobile est divisée en cellules se partageant l'ensemble de la bande de fréquences suivant deux méthodes : FCA et DCA [KAT 96]. Dans FCA, à chacune des cellules lui est allouée un ensemble fixe de canaux et un canal est alloué à chaque usager. Nous rappelons également que l'un des soucis majeurs lors de la conception des réseaux

mobiles est la réduction de la probabilité de coupure lors d'un transfert inter-cellulaire.

Les techniques de réservation de canaux (ou *Guard Channel*) permettent de réduire la probabilité de coupure de communication en réservant, dans chaque cellule, des canaux à l'usage exclusif des handoffs [ZHA 89], [KAT 96]. Si ces techniques sont simples à dimensionner lorsqu'on considère une seule classe de trafic (i.e. appel téléphonique), elles deviennent beaucoup plus compliquées à optimiser, et moins optimales dans un contexte de trafic multi-classes.

En effet, dans un contexte de trafic multi-classes, il est parfois préférable de bloquer un appel d'une classe moins prioritaire, et d'accepter un autre appel appartenant à une classe de priorité supérieure. Le contrôle d'admission des appels présenté dans cette section, permet ce genre de mécanismes. Il est obtenu en utilisant l'algorithme d'apprentissage par renforcement : Q-learning présenté dans la section précédente.

### 11.3.1. *Formulation du problème*

Concentrons-nous sur une simple cellule FCA avec $N$ canaux disponibles, et deux[2] classes de trafic *C1* et *C2*. Les appels de la première classe demandent un seul canal, alors que les appels de la seconde demandent deux canaux. Ce système cellulaire peut être considéré comme un système à événement discret. Les principaux événements pouvant se produire dans une cellule sont les arrivées et les départs d'appels. Ces événements sont modélisés par des variables stochastiques avec des distributions appropriées. En particulier, les arrivées des nouveaux appels et des handoffs sont Poissonniennes. Le temps de séjour de chaque appel est exponentiellement distribué. Les appels arrivent dans la cellule puis partent, soit en effectuant un handoff vers une autre cellule, soit en se terminant normalement. Le réseau devra alors choisir d'accepter ou de rejeter ces demandes de connexion. En retour, il collecte l'ensemble des revenus reçus de la part des clients acceptés (les gains), ainsi que l'ensemble des revenus reçus de la part des clients rejetés (les pertes). L'objectif de l'opérateur du réseau est de trouver une politique de CAC qui maximise l'ensemble des gains à long terme et qui réduise les probabilités de blocage des handoffs ainsi que l'ensemble des pertes.

Nous avons choisi la description de l'ensemble des états comme étant $s=((x_1,x_2),e)$, où $x_i$ est le nombre d'appels en cours de la classe $C_i$, et $e$ représente l'arrivée d'un nouvel appel ou d'un appel de handoff dans la cellule. Quand un

2. L'idée peut facilement être étendue à plusieurs classes de trafic.

événement se produit, l'agent doit choisir une des actions possibles $A(s)$={rejeter, accepter}. À la fin d'un appel, aucune mesure ne doit être prise.

À chaque type d'appel est associé un revenu. Étant donné que l'objectif principal de l'opérateur du réseau est de diminuer les probabilités de blocage des handoffs, des valeurs de revenus assez grandes leur ont été affectées. Les valeurs de revenus pour les appels de la classe *C1* sont plus importantes que celles des appels de la classe *C2*, étant donné que *C1* est supposée prioritaire par rapport à *C2*.

L'agent a la tâche de déterminer une politique d'acceptation des appels en connaissant uniquement l'état du réseau. Ce système constitue un SMDP avec comme espace d'états un ensemble fini $S = \{s=(x, e)\}$ et comme espace d'actions possibles l'ensemble fini $A=\{0,1\}$ dont le Q-learning est la solution idéale.

### 11.3.2. *Implantation de l'algorithme*

Après avoir formulé le problème de contrôle d'admission des appels sous forme d'un SMDP, nous allons décrire deux implantations de l'algorithme Q-learning capables de le résoudre. Nous les avons nommées TRL-CAC (*Table Reinforcement Learning CAC*) et NRL-CAC (*Neural network Reinforcement Learning CAC*). TRL-CAC utilise une simple table pour représenter la fonction $Q$ (l'ensemble des Q-valeurs). En revanche, NRL-CAC utilise un réseau de neurones multicouches [MIT 97]. Les différences entre ces deux algorithmes (TRL-CAC et NRL-CAC) s'expriment en terme de complexité de calcul et de taille mémoire.

L'approche utilisant une table, TRL-CAC, est la méthode la plus simple et la plus efficace. Cette approche conduit à un calcul exact et elle est complètement conforme aux hypothèses de structures faites afin de prouver la convergence de l'algorithme du Q-learning. Cependant, lorsque l'ensemble des paires état-action $(s,a)$ est grand ou lorsque les variables d'entrée $(x,e)$ constituant l'état $s$ sont des variables continues, l'utilisation d'une simple table devient rédhibitoire à cause des immenses besoins de stockage. Dans ce cas, quelques fonctions d'approximation, telles que l'agrégation des états, les réseaux de neurones [MIT 97], [MCC 43] ou même les arbres de régression [BRE 84] peuvent être utilisés de manière efficace. Le réseau de neurones utilisé dans NRL-CAC est constitué de 4 entrées, 10 unités cachées et une unité de sortie.

#### 11.3.2.1. *Implantation*

Lorsqu'un appel arrive (nouvel appel ou handoff), l'algorithme détermine si la qualité de service n'est pas violée en acceptant cet appel (en vérifiant tout

simplement s'il y a assez de canaux disponibles dans la cellule). Si cette qualité de service est violée, l'appel est rejeté ; sinon l'action est choisie selon la formule :

$$a = \arg\max_{a \in A(s)} Q^*(s,a) \qquad [11.4]$$

où $A(s)$={1=accepter, 0=rejeter}.

En particulier, [11.4] implique les procédures suivantes : lorsqu'un appel arrive, la Q-valeur d'acceptation ainsi que la Q-valeur de rejet de l'appel sont déterminées soit à partir de la table (TRL-CAC), soit à partir du réseau de neurones (NRL-CAC). Si le rejet a une plus grande valeur, l'appel est alors rejeté. Dans le cas contraire, l'appel est accepté.

Dans ces deux cas, et pour apprendre les valeurs optimales $Q^*(s,a)$, la fonction est mise à jour à chaque transition du système d'un état $s$ vers un état $s'$. Pour les deux algorithmes, ceci se fait de la façon suivante :

– TRL-CAC : c'est la formule [11.2] qui est utilisée pour mettre à jour la Q-valeur appropriée dans la table ;

– NRL-CAC : quand un réseau de neurones est utilisé pour stocker la fonction $Q$, une seconde procédure d'apprentissage est nécessaire pour apprendre les poids du réseau de neurones. Cette procédure utilise l'algorithme de rétro-propagation (*BP – Back Propagation*) [MIT 97]. Dans ce cas, $\Delta Q$ définie par la formule [11.3] est employée comme signal d'erreur qui est rétro-propagé dans les différentes couches du réseau de neurones.

11.3.2.2. *Exploration*

Pour une exécution correcte et efficace de l'algorithme du Q-learning, toutes les paires potentiellement importantes des état-action $(s,a)$ doivent être explorées. Pour cela, pendant une période d'apprentissage assez longue, l'action est choisie non selon la formule [11.4], mais selon la formule [11.5] suivante avec une probabilité d'exploration $\in$ :

$$a = \arg\min_{a \in A(s)} visites(s,a) \qquad [11.5]$$

où visites$(s,a)$ désigne le nombre de fois où la configuration $(s,a)$ a été visitée. Cette heuristique, appelée $\in$-gloutonne (ou $\in$-*directed*), accélère significativement la convergence des Q-valeurs. Ces valeurs sont, dans un premier temps, calculées en utilisant cette heuristique pendant une période d'apprentissage. Dans un second temps, ces valeurs seront employées pour initialiser les Q-valeurs dans les deux algorithmes de CAC.

### 11.3.3. *Résultats expérimentaux*

Afin d'évaluer les avantages de nos algorithmes, nous avons utilisé une simulation à événements discrets pour représenter le réseau cellulaire. La cellule FCA possède $N = 24$ canaux. La constante de propagation temporelle $\gamma$ a été fixée à 0.5 et la probabilité d'exploration $\in$ à 1. Les critères de performance utilisés afin de comparer ces algorithmes sont : (i) les gains, (ii) les pertes et (iii) les probabilités de coupure. Les gains représentent la somme des revenus dus à l'acceptation de nouveaux appels ou de handoffs pour les deux classes de trafic dans la cellule. Quant aux pertes, elles représentent la somme des revenus dus au rejet de nouveaux appels ou à l'échec d'un transfert inter-cellulaire. Les probabilités de coupure de tous les appels confondus (*C1* et *C2*) ont été calculées en utilisant la formule [11.6] suivante :

$$P_{HO} = \frac{\textit{nombre d'échecs de handoff dans le système}}{\textit{nombre de tentatives de handoff dans le système}} \qquad [11.6]$$

Quant aux probabilités de coupure de chacune des classes de trafic, elles ont été calculées pour chaque classe de trafic $C_i$ en utilisant la formule [11.7] suivante :

$$P_{HO(C_i)} = \frac{\textit{nombre d'échecs de handoff de type } C_i \textit{ dans le système}}{\textit{nombre de tentatives de hondoff de type } C_i \textit{ dans le système}} \qquad [11.7]$$

Nous avons comparé nos politiques à celle que nous avons appelée politique gloutonne (ou *greedy*) [TON 00] (politique qui accepte un nouvel appel ou un appel de handoff si la contrainte de capacité n'est pas violée en acceptant cet appel). Nous les avons également comparé au mécanisme de réservation des canaux (ou « *guard channel* »). Le mécanisme de réservation de canaux permet de partager la capacité de la cellule entre les nouveaux appels et les appels de handoffs, en donnant une certaine priorité à ces derniers. Ceci est réalisé en réservant, dans chaque cellule, des canaux à l'usage exclusif des handoffs (canaux de garde). Le nombre de canaux réservés pour les appels de handoffs est donc un paramètre de grande importance dans ce genre de mécanisme. Dans un contexte multi-service, la question qui se pose est la suivante : combien de canaux peuvent être réservés aux handoffs pour chacune des classes de trafic, afin de maximiser l'ensemble des revenus ? Pour répondre à cette question, nous avons développé un modèle mathématique classique où une cellule est représentée par une file multi-serveur classique de type *M*/*M*/*N*/*N* et où chacun des serveurs représente un canal de communication.

Nous avons effectué un ensemble de simulations, incluant : (i) le cas d'une charge de trafic constante pour toutes les classes de trafic, (ii) le cas de charges de trafic variables, et (iii) le cas d'une charge de trafic variable dans le temps. Nous

n'allons pas présenter tous les résultas obtenus, mais les lecteurs intéressés peuvent consulter les références [SEN1 03], [SEN2 03] pour plus de détails.

– Charge de trafic constante : la première expérience considère une charge de trafic constante pour les deux classes C1 et C2. Les paramètres de simulation utilisés sont les mêmes que ceux utilisés pendant la période d'apprentissage ;

– Charges de trafic variables : dans cette deuxième expérience, nous avons utilisé les mêmes politiques que dans l'expérience précédente (charge de trafic constante), mais avec six charges de trafic différentes (pour les deux classes C1 et C2) ;

– Charge de trafic variable dans le temps : dans cette dernière expérience, nous avons utilisé encore une fois les mêmes politiques que dans la première expérience, mais avec une charge de trafic variable dans le temps. En effet, la charge de trafic dans un système cellulaire est variable dans une même journée. Nous avons utilisé le modèle de trafic présentée dans le Figure 11.3 et décrivant la variation des taux d'arrivée pendant une journée ouvrable. Les heures de pointe apparaissent à 11 heures et à 16 heures.

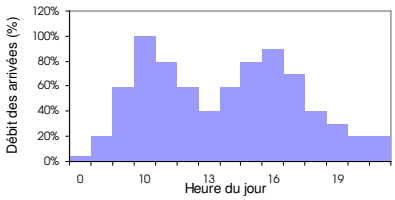


**Figure 11.3.** *Modèle de trafic d'une journée ouvrable.*

Pour toutes ces expériences, l'ensemble des résultats a bien confirmé que les algorithmes proposés sont plus performants que les autres heuristiques et ceci pour toutes les charges de trafic, et plus particulièrement quand la charge de trafic est grande. Par exemple, la Figure 11.4 donne les résultats de simulation, avec l'hypothèse que les deux classes de trafic suivent le même modèle de trafic. Les probabilités de blocage ont été calculées sur une base d'une heure. Nous avons comparé nos algorithmes à : (1) guard channel avec seuils fixes – ces seuils ont été calculés pour une charge de trafic constante donnée ; et (2) guard channel avec seuils optimisés – ces seuils ont été calculés pour chaque valeur de la charge de trafic. Les améliorations des algorithmes proposés sont bien apparentes par rapport à la politique gloutonne, et particulièrement pendant les pics de trafic (vers 11 heures et vers 16 heures). Nous pouvons remarquer que la méthode de réservation de canaux avec seuils optimisés donne de meilleurs résultats que les deux algorithmes basés sur le Q-learning. En revanche, cette méthode suppose que les seuils optimisés soient calculés, hors ligne, pour chaque valeur de la charge de trafic. Au contraire, les algorithmes de contrôle d'admission basés sur le Q-learning (TRL-CAC et NRL-

CAC) possèdent des capacités d'adaptation et de généralisation, et donc les valeurs optimales des Q-valeurs ne sont pas recalculées pour chaque charge de trafic.

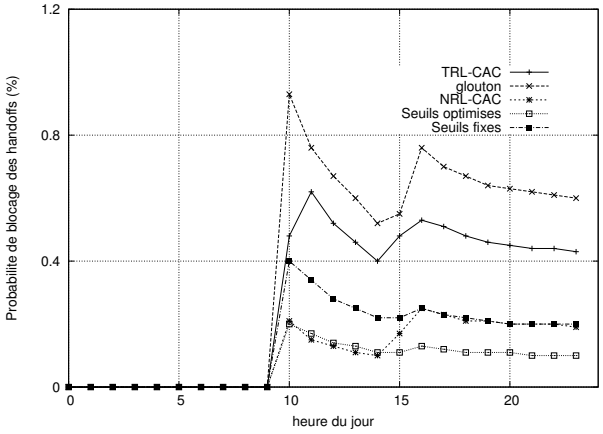


**Figure 11.4.** *Probabilité de blocage des handoffs avec une charge de trafic variable dans le temps.*

## 11.4. Allocation dynamique des ressources

Dans cette section nous allons présenter un nouveau mécanisme permettant de résoudre le problème de l'allocation dynamique des ressources, considérant également le contrôle d'admission d'appels dans les systèmes DCA supportant plusieurs classes de trafic.

Rappelons que dans les stratégies d'allocation dynamique des ressources (DCA), tous les canaux disponibles dans le système sont mis dans un pool commun (ou *common pool*) utilisable par toutes les stations de base [KAT 96]. Lors d'une demande de communication, une cellule choisit un canal du pool commun[3], qui sera restitué à la fin de l'appel.

Différents algorithmes d'allocation dynamique de ressources ont été comparés, en termes de performance, flexibilité et complexité. L'une des stratégies la plus utilisée dans la littérature appartient à une classe d'algorithmes appelée « DCA à recherche exhaustive - *exhaustive searching DCA* » [NIE 99], [ZHA 89], [COX 72], [DEL 93], [DIM 93], [SIV 90]. Dans ce type d'algorithmes, une récompense (coût) est associée à chaque canal disponible. Quand un nouvel appel arrive, le système recherche exhaustivement le canal avec la récompense maximale (coût minimal) et

3. Tout en respectant le rapport signal/interférence *C/I*.

l'affecte à l'appel. La majorité de ces propositions ne considèrent pas les politiques de contrôle d'admission CAC en tant que moyens pour empêcher la congestion [LEV 97]. Le rapport entre l'allocation de ressources et le contrôle d'admission ont été étudiés précédemment [NAG 95], [TAJ 88], [YAN 94].

Cette section présente une nouvelle approche pour résoudre le problème de l'allocation dynamique des ressources, considérant également le contrôle d'admission d'appels dans les systèmes DCA. Pour les raisons citées dans la section précédente, le contrôle d'admission d'appels est indispensable lorsque le réseau supporte plusieurs classes de clients. Les politiques d'allocation sont obtenues en utilisant l'algorithme d'apprentissage par renforcement Q-learning. Les fonctions principales du mécanisme proposé, appelé Q-DCA et présenté dans cette section, sont : (i) accepter des clients et en rejeter d'autres, et (ii) allouer le meilleur canal disponible aux clients acceptés. L'objectif consiste à maximiser le cumul des revenus reçus au cours du temps.

Nous allons maintenant décrire brièvement la formulation du problème par un SMDP ainsi que l'implantation de l'algorithme du Q-learning capable de résoudre ce SMDP.

### 11.4.1. *Formulation du problème*

Cette contribution est une extension du travail de *Nie* et *Haykin* [NIE 99] et faisant partie des algorithmes DCA à recherche exhaustive. Nous envisageons de résoudre non seulement le problème d'allocation dynamique de ressources, mais également le problème du contrôle d'admission des appels dans un réseau cellulaire. Ce réseau contient $N$ cellules et $M$ canaux disponibles maintenus dans un pool commun. Il supporte deux classes de trafic (contrairement à [NIE 99] où une seule classe de trafic –appel téléphonique– était prise en compte). Chaque canal peut être temporairement alloué à n'importe quelle cellule, à condition que la contrainte sur la distance de réutilisation soit satisfaite (une qualité donnée de signal soit maintenue).

Les appels arrivent dans la cellule, puis partent selon des distributions appropriées. Le réseau devra donc choisir d'accepter ou de rejeter ces demandes de connexion. Si l'appel est accepté, le système lui alloue un des canaux disponibles du pool commun. L'objectif pour l'opérateur du réseau est de trouver une politique d'allocation dynamique de ressources capable de maximiser l'ensemble des gains à long terme et de réduire les probabilités de blocage de handoff (contrairement à [NIE 99] qui n'accorde aucune priorité aux appels de handoff).

Nous avons choisi l'ensemble des états comme étant $\{s=(i,D(i),(x_1,x_2),e_i)\}$, où $D(i)$ représente le nombre de canaux disponibles dans la cellule $i$ où s'est produit

l'événement $e_i$, $x_k$ représente le nombre d'appels en cours de la classe $C_k$, et $e_i$ indique l'arrivée d'un nouvel appel ou d'un appel de handoff dans une cellule *i*. Quand un événement se produit, l'agent doit choisir une des actions possibles $A(s)=\{0=\text{rejeter}\}\cup\{1,\ldots,M\}$. Quand un appel se termine, aucune mesure ne doit être prise. L'agent aura à déterminer les politiques permettant d'accepter ou de rejeter un appel et à lui allouer, dans le cas d'acceptation, le canal qui permettra de maximiser le cumul des gains reçus, en ne connaissant que l'état actuel du réseau *s*. Ce système constitue un SMDP avec comme espace d'états un ensemble fini $S = \{(i, D(i), x, e)\}$, et comme espace d'actions possibles un ensemble fini $A=\{0,1,\ldots,M\}$. Le choix des revenus que nous avons utilisé tient en compte des interférences inter-cellulaires et a comme conséquence une situation dans laquelle les canaux déjà utilisés dans les cellules compactes[4] [ZHA 91] auront plus de chance à être choisis. Contrairement à d'autres travaux, Q-DCA prend en considération le type de l'appel et accorde, ainsi, une priorité aux appels de handoff.

### 11.4.2. *Implantation de l'algorithme*

Après avoir formulé le problème sous forme de SMDP, nous allons décrire l'implantation de l'algorithme Q-learning capable de le résoudre. Le système cellulaire étudié est composé de $N = 36$ cellules hexagonales et $M = 70$ canaux disponibles dans un pool commun. Nous prenons une distance de réutilisation $D = \sqrt{21}R$ ($R$ représente le rayon de la cellule). Ceci implique que si un canal est alloué à une cellule *i*, alors il ne peut pas être réutilisé dans les deux rangées adjacentes à *i* en raison des interférences co-canaux. Ainsi, il y a au plus 18 cellules interférentes avec chaque cellule du système. La constante de propagation temporelle $\gamma$ a été fixée à 0.5, et le taux d'apprentissage $\alpha$ à 0.1.

Dans la section précédente, nous avons utilisé un réseau de neurones pour représenter les Q-valeurs (NRL-CAC), mais cette fois-ci nous avons choisi une approximation à base d'agrégation d'états. Ainsi, au lieu de spécifier précisément le nombre d'appels $x_i$ pour chaque classe de trafic $C_i$, nous avons choisi de caractériser le trafic comme suit : (1) bas, (2) moyen, (3) élevé. L'espace des états est ainsi réduit ; et une simple table peut être utilisée pour représenter les états agrégés. Etant donné que des états identiques (i.e. états ayant un nombre d'appels en cours similaire) ont les mêmes Q-valeurs, la perte en performance liée à l'agrégation devient négligeable [TON 99].

---

4. Les cellules compactes sont les cellules avec une distance moyenne minimum entre les cellules co-canaux.

11.4.2.1. *Implantation*

Lorsqu'un appel arrive (nouvel appel ou appel de handoff) dans la cellule *i*, l'algorithme détermine si la qualité de service n'est pas violée en acceptant cet appel (en vérifiant tout simplement s'il y a des canaux libres dans le pool commun). Si cette qualité de service est violée, l'appel est rejeté ; sinon l'action est choisie selon la formule [8] suivante :

$$a = \arg \max_{a \in A(s)} Q^*(s, a) \qquad [8]$$

où $A(s)$={0=rejeter, 1,2,…,$M$}.

La formule [8] implique les procédures suivantes. Quand il y a une tentative de connexion d'un appel dans la cellule *i*, la Q-valeur de rejet ($a$ = 0) ainsi que les Q-valeurs d'acceptation ($a$ = 1,2,…,$M$) sont déterminées à partir de la table des Q-valeurs. Les Q-valeurs d'acceptation incluent les différentes Q-valeurs correspondant aux choix de chacun des canaux $a$ ($a$ = 1,2,…,$M$), pour servir l'appel. Si le rejet a la plus grande valeur, l'appel est alors rejeté. Dans le cas contraire, si l'une des valeurs d'acceptation a la plus grande valeur, l'appel est accepté et le canal $a$ lui est alloué.

11.4.2.2. *Exploration*

Pour une exécution correcte et efficace de l'algorithme Q-DCA, l'action est choisie non selon la formule [11.4], mais selon une distribution de Boltzmann [WAT 89], ceci pendant une période d'apprentissage assez longue. L'idée est de favoriser, au début, l'exploration (la probabilité d'exécuter des actions autres que celles avec la Q-valeur la plus élevée), en utilisant toutes les actions possibles avec une même probabilité. L'idée ensuite est de tendre, au fur et à mesure, vers l'utilisation de l'action de plus grande Q-valeur. Les valeurs apprises seront employées après pour initialiser les Q-valeurs dans Q-DCA.

### 11.4.3. *Résultats expérimentaux*

Afin d'étudier les performances de Q-DCA, un ensemble de simulations a été réalisé. Nous avons comparé Q-DCA à la politique glouton-DCA[5] [TON 99], ainsi qu'à l'algorithme DCA-Nie [NIE 99]. Les performances des algorithmes ont été également évaluées en terme de gains, de pertes, ainsi que de probabilités de blocage

5. Glouton-DAC: Politique qui choisit aléatoirement un canal pour servir un appel sans aucune mesure d'interférences. Chacun des $M$ canaux a la même probabilité d'être choisi pour servir le nouvel appel.

des handoffs. Un ensemble de simulations a été effectué, incluant : (i) le cas d'une charge de trafic uniformément répartie sur toutes les cellules, (ii) le cas d'une charge non-uniformément répartie, (iii) le cas d'une charge de trafic variable dans le temps, et (iv) le cas d'une panne d'équipements. Nous n'allons pas présenter tous les résultas obtenus, mais les lecteurs intéressés peuvent consulter les références [SEN1 03], [SEN3 03] pour plus de détails.

– Répartition uniforme du trafic : la première expérience considère une charge de trafic constante dans les 36 cellules pour les deux classes de trafic. Nous avons utilisé les politiques apprises pendant la période d'apprentissage, mais avec cinq charges de trafic différentes (pour les deux classes C1 et C2) ;

– Répartition non uniforme du trafic : dans cette deuxième expérience, nous avons utilisé les politiques apprises pendant la période d'apprentissage mais les charges de trafic dans ce cas ne sont plus uniformément réparties sur les 36 cellules. La charge de trafic moyenne considérée était de 7.5 Erlangs ;

– Charge de trafic variable dans le temps : dans la troisième expérience, nous avons voulu tester les performances de Q-DCA lorsque la charge du trafic change dans le temps ;

– Panne d'équipement dans un système DCA : dans la dernière expérience, nous avons simulé un échec d'équipement par le fait que quelques canaux deviennent temporairement indisponibles. Au début de la simulation, il y a 70 canaux disponibles dans le système. Mais entre 10 heures et 15 heures, nous avons temporairement suspendu 0, 3, 5 et 7 canaux.

Pour toutes ces expériences, l'ensemble de résultats montre les possibilités qu'offre l'apprentissage par renforcement afin d'apprendre la meilleure politique d'admission et d'allocation dynamique de ressources. Les résultats des politiques utilisant Q-DCA indiquent des améliorations significatives par rapport aux autres politiques. Ces amélioration sont également légèrement meilleures que celles de DCA-Nie. Ainsi, Q-DCA possède une certaine aptitude à généraliser et à s'adapter aux changements dans les conditions de trafic. Par exemple, la Figure 11.5 montre l'impact de l'échec des canaux sur les probabilités de coupure en utilisant l'algorithme Q-DCA. Nous remarquons que Q-DCA possède une certaine robustesse contre les situations de panne d'équipements et s'adapte facilement surtout dans le cas où 3/5 canaux ont été suspendus.

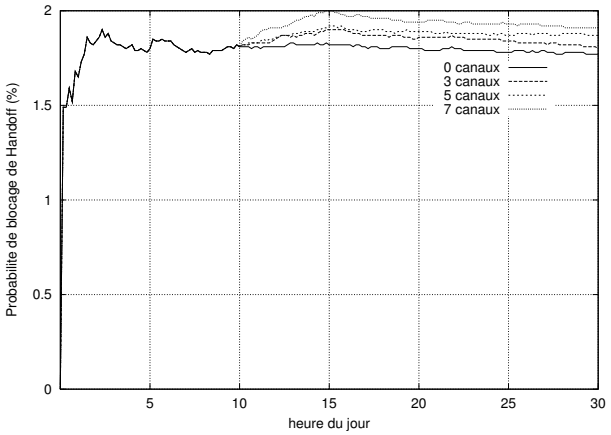


**Figure 11.5.** *Performances de Q-DCA lors de pannes de canaux.*

## 11.5. Conclusion

Les premiers réseaux cellulaires ont été conçus pour offrir principalement un service de téléphonie. Les systèmes cellulaires actuels promettent, quant à eux, une diversification des services proposés avec un débit nettement supérieur. La diversification des services (voix, messages courts, services multimédias, accès au réseau Internet, etc.) impose plusieurs niveaux de qualité de service (QoS) à garantir. Cependant, la disponibilité des ressources radios durant toute la durée de l'appel n'est pas nécessairement garantie, et les utilisateurs mobiles peuvent ainsi subir une dégradation/coupure du service. Partant du fait que la coupure/dégradation d'un appel en cours, appartenant à une classe de service de haute priorité, est généralement moins désirée que la coupure/échec de connexion d'un appel appartenant à une classe de priorité inférieure, de nouveaux mécanismes de contrôle d'admission d'appels (CAC) sont fortement nécessaires. En effet, un contrôle d'admission d'appel efficace est exigé pour remédier à cette limitation du nombre de ressources radio disponibles sur l'interface radio des réseaux cellulaires. Une gestion efficace des ressources radio, par l'intermédiaire de politiques d'allocation dynamique, s'avère également indispensable pour remédier à ce genre de problèmes. Ces nouveaux mécanismes consistent à définir des règles de gestion des ressources pour chaque classe de trafic, dans le but d'optimiser le taux d'utilisation et de satisfaire de multiples contraintes de qualité de service.

Afin de remédier à ce genre de problèmes, de nombreuses propositions existent dans la littérature ; leur objectif principal est d'éviter aux usagers le désagrément causé par des coupures de communication. Nous avons cependant constaté que ces

solutions ignorent souvent l'expérience et la connaissance qui pourraient être acquises pendant l'exécution en temps réel du système.

Les contributions présentées dans ce chapitre, relatives aux réseaux cellulaires, ont pour principal objectif de tirer profit de cette expérience et de cette connaissance afin d'optimiser certains problèmes rencontrés dans les réseaux cellulaires. Dans la première contribution il fallait trouver une nouvelle approche pour résoudre le problème de contrôle d'admission d'appel CAC dans un réseau cellulaire multi-services où les canaux sont alloués de façon permanente aux cellules. Quant à la seconde, il fallait trouver une nouvelle approche au problème d'allocation dynamique des ressources dans un réseau cellulaire multi-services. Cette dernière est originale, puisqu'elle combine à la fois la recherche de la politique de CAC optimale et de la meilleure stratégie d'allocation dynamique de canaux. Ces mécanismes proposés afin de résoudre des problématiques aussi complexes que celles liées aux réseaux cellulaires utilisent l'apprentissage par renforcement comme solution. Ce sont des solutions originales et de nature intelligente. Outre l'originalité de ces mécanismes, les avantages acquis, en utilisant de telles approches, peuvent être récapitulés comme suit. Contrairement à d'autres travaux (études basées sur des modèles mathématiques ou sur des simulations, supposant des paramètres expérimentaux fixes), ces solutions sont adaptables aux variations de l'état du réseau (i.e. variations des conditions de trafic, pannes d'équipements, etc.). De part leurs natures distribuées, elles sont facilement implantables dans chaque station de base, ce qui les rend plus attrayantes. Les tâches de contrôle d'admission et d'allocation dynamique des canaux sont déterminées rapidement et avec peu d'efforts de calcul. Elles sont obtenues grâce à une simple spécification de préférences entre les classes de service. Nous avons démontré également, par un large ensemble d'expériences, que ces mécanismes donnent de meilleurs résultats comparés à d'autres heuristiques. Ce sont des algorithmes distribués, et les informations de signalisation échangées entre les stations de base sont quasi nulles. Ces mécanismes sont donc plus attrayants, en raison de la simplicité de leur mise en œuvre.

En conclusion générale de ce chapitre, nous pouvons dire que notre principale contribution a été de proposer et d'éprouver des mécanismes permettant de résoudre un certain nombre de problèmes rencontrés dans les réseaux mobiles (contrôle d'admission des appels et allocation dynamique de ressources). Nous avons pu démontrer qu'il est possible d'utiliser des techniques venant du monde de l'Intelligence Artificielle (IA), et plus spécialement des techniques d'apprentissage afin d'élaborer des mécanismes très efficaces, robustes et très simples à mettre en œuvre.

## 11.6. Bibliographie

[AGH 01] K. AL AGHA, G.PUJOLLE, ET G.VIVIER, « Réseaux de mobiles et réseaux sans fil », Editions Eyrolles, 2001.

[BOY 94] J. A. BOYAN AND M. L. LITTMAN, « Packet routing in dynamically changing networks: A reinforcement approach », *In Advances in Neural Information Processing Systems (NIPS'94)*, volume 6, pages 671-678, San Mateo, CA, 1994.

[BRE 84] L. BREIMAN, J.H. FRIEDMAN, R.A. OLSEN, AND C.J. STONE, « *Classification and Regression Trees* », Chapman & Hall, 1984.

[CAL 92] G. CALHOUN, « Radio Cellulaire numérique", TEC and Doc, 1992.

[COR] M.-M. CORSINI, *cours sur l'apprentissage par renforcement*, http://www.sm.u-bordeaux2.fr/~corsini/Cours/HeVeA/rl.html.

[COX 72] D. C. COX AND D. O REUDINK, « Dynamic channel assignment in two dimensional large mobile radio systems », Bell Syst. Tech. J., vol. 51, pp. 1611-1627, 1972.

[DEL 93] E. DEL RE, R. FANTACCI, AND L. RONGA, « A dynamic channel allocation technique based on Hopfield neural networks », *IEEE Trans. Vehicular Technology*, vol. 45, pp. 26-32, février 1996.

[DIM 93] D. D. DIMITRIJEVIC AND J. VUCETIC, « Design and performance analysis of the algorithms for channel allocation in cellular networks », *IEEE Trans. Vehicular Technology*, vol. 42, pp. 526-534, novembre 1993.

[GIB 96] J. D. GIBSON, « The telecommunications Handbook », *IEEE press*, 1996.

[KAT 96] I. KATZELA, M. NAGHSHINEH, « Channel assignement schemes for cellular mobile telecommunications systems », *IEEE Personal Communications Magazine*, juin 1996.

[LEV 97] D. A. LEVINE, I. F. AKYILDIZ AND M. NAGHSHINEH, « A Resource Estimation and Call Adaptation Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept », *IEEE/ACM Transactions on Networking*, volume 5, n° 1, pp. l-12, février 1997.

[MANET] IETF MANET Working Group (Mobile Ad hoc NETworks), www.ietf.ora/html.charters/manet-charter.html.

[MAR 97] P. MARBACH, J. N. TSITSIKLIS, « A Neuro-Dynamic Approach to Admission Control in ATM Networks: The Single Link Case », *ICASSP'97*, 1997.

[MAR 98] P. MARBACH, O. MIHATSCH, M. SCHULTE AND J. N. TSITSIKLIS, « Reinforcement learning for call admission control and routing in integrated service networks », in Jordan, M., et al., ed. Advances in NIPS 10, MIT Press, 1998.

[MAR 00] P. MARBACH, O. MIHATSCH AND J. N. TSITSIKILS, « Call admission control and routing in integrated services networks using neuro-dynamic programming », *IEEE Journal on Selected Areas in Communications (JSAC'2000)*, vol. 18, n°. 2, pp. 197 –208, février 2000.

[MCC 43] W.S. MCCULLOCH ET W. PITTS, « A logical calculus of the ideas Imminent in Nervous Activity », *Bulletin of math. Biophysics*, vol. 5, 1943.

[MIT 97] T. M. MITCHELL, « *Machine Learning* », McGraw-Hill companies, Inc., 1997.

[MIT 98] MITRA, M. I. REIMAN AND J. WANG, « Robust dynamic admission control for unified cell and call QoS in statistical multiplexers », *IEEE Journal on Selected Areas in Communications (JSAC'1998)*, vol. 16, no. 5, pp. 692-707, juin 1998.

[NAG 95] M. NAGHSHINEH, O. SCHWARTZ, « Distributed call admission control in mobile/wireless networks », *PIMRS, Proceedings of Personal Indoor and mobile radio communications*, 1995.

[NIE 99] J. NIE AND S. HAYKIN, « A Q-Learning based dynamic channel assignment technique for mobile communication systems », *IEEE Transactions on Vehicular Technology*, vol. 48, n°. 5, septembre 1999.

[RAM 96] R. RAMJEE, R. NAGARAJAN AND D. TOWSLEY, « On Optimal Call Admission Control in Cellular Networks », *IEEE INFOCOM*, pp. 43-50, San Francisco, CA, mars1996.

[SIV 90] K. N. SIVARAJAN, R.J MCELIECE, AND J.W.KETCHUM, « Dynamic channel assignment in cellular radio », *Proc. IEEE 40th Vehicular Technology Conf.*, pp. 631-637, mai 1990.

[SEN1 03] S. SENOUCI, « Application de techniques d'apprentissage dans les réseaux mobiles », Thèse de Doctorat de l'Université de Pierre et Marie Curie, Paris, octobre 2003.

[SEN2 03] S. SENOUCI, A.-L. BEYLOT, G. PUJOLLE, « Call Admission Control in Cellular Networks: A Reinforcement Learning Solution », *ACM/Wiley International Journal of Network Management*, vol. 14, n° 2, mars-avril 2003.

[SEN3 03] S. SENOUCI, AND G. PUJOLLE, « New Channel Assignments in Cellular Networks: A reinforcement Learning Solution », *in Asian Journal of Information Technology (AJIT'2003)*, pp. 135-149, vol. 2, n° 3, juillet-septembre, 2003, Grace Publications Network.

[SRI 00] M. SRIDHARAN, G. TESAURO, « Multi-agent Q-learning and Regression Trees for Automated Pricing Decisions », *in Proceedings of the Seventeenth International Conference on Machine Learning (ICML'00)*, Stanford, CA, juin-juillet, 2000.

[SUT 98] R. S. SUTTON AND G. BARTO, ANDREW, « Reinforcement Learning: An Introduction », *MIT Press*, 1998.

[TAJ 88] J. TAJIMA AND K. IMAMURA, « A strategy for exible channel assignment in mobile communication systems », *IEEE Transaction on Vehicular Technology*, vol. 37, pp. 92–103, mai 1988.

[TON 99] H. TONG, « Adaptive Admission Control for Broadband Communications », Ph.D. thesis, University of Colorado, Boulder, été 1999.

[TON 00] H. TONG AND T. X. BROWN, « Adaptive Call Admission Control under Quality of Service Constraint: a Reinforcement Learning Solution », *IEEE Journal on Selected Areas in Communications (JSAC'2000)*, vol. 18, n°. 2, pp. 209-221, février 2000.

[WAT 89] C. J. C. H. WATKINS, « Learning from delayed rewards », PhD. thesis, University of Cambridge, Psychology Department, 1989.

[WAT 92] C. J. C. H. WATKINS AND P. DAYAN, « Q-learning », *Machine Learning*, vol. 8, pp. 279-292, 1992.

[YAN 94] W.-B. YANG AND E. GERANIOTIS, « Admission policies for integrated voice and data traffic in CDMA packet radio networks », *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 654–664, mai 1994.

[ZHA 89] M. ZHANG AND T.S. YUM, « Comparisons of channel assignment strategies in cellular mobile systems », *IEEE Trans. Vehicular Technology*, vol. 38, n° 1, pp. 211-215, juin 1989.

[ZHA 91] M. ZHANG AND T.S. YUM, « The nonuniform compact pattern allocation algorithm for cellular mobile systems », *IEEE Trans. Vehicular Technology*, vol. 40, n° 2, pp. 387-391, mai 1991.