Apprentissage pour l'Accès Opportuniste au Spectre : Prise en Compte des Erreurs d'Observation

Wassim JOUINI, Christophe MOY, Jacques PALICOT,
SCEE/IETR
SUPELEC, Avenue de la Boulaie, CS 47601
35576 Cesson Sévigné Cedex, France.
wassim.jouini@supelec.fr, christophe.moy@supelec.fr, jacques.palicot@supelec.fr

Résumé – Le contexte considéré est celui de l'Accès Opportuniste au Spectre. On suppose qu'il existe un utilisateur dit "secondaire" intéressé par une bande de fréquence dédiée à un réseau d'utilisateurs prioritaires dits "primaires". Les capacités d'écoute et d'exploitation du spectre de l'utilisateur secondaire sont supposées limitées à une sous-bande de la bande de fréquences sondée. Dans une précédente étude, les auteurs avaient suggéré une classe d'algorithme d'apprentissage par renforcement, connue sous le nom d'Upper Confidence Bound Algorithms, pour répondre à cette problématique. Le mécanisme d'apprentissage repose essentiellement sur l'observation de l'activité des utilisateurs primaires dans les bandes spectrales sollicitées par l'utilisateur secondaire. L'objectif de ce papier est d'évaluer l'impact des erreurs d'observations de la présence ou absence des utilisateurs primaires sur les performances d'apprentissage de l'utilisateur secondaire. Ces travaux montrent que la convergence vers le canal dit "optimal", i.e., le plus libre en probabilité, reste rapide malgré les erreurs d'observation liées à la détection des utilisateurs primaires. Ce résultat est mis en avant à travers un théorème qui montre que le temps passé par l'algorithme à observer un canal sous-optimal est borné par une fonction logarithmique du temps. Ainsi, les pertes de performances dues aux erreurs d'observations sont quantifiées en fonction des caractéristiques du capteur (probabilité de fausses alarmes notamment). Ces résultats sont illustrés à l'aide de simulations.

Abstract – In this paper we consider the problem of exploiting spectrum resources within the Opportunistic Spectrum Access context. We mainly focus on the case where one secondary user (SU) probes a pool of possibly available channels dedicated to a primary network. The SU is assumed to have imperfect sensing abilities. We, first, model the problem as a Multi-Armed Bandit problem with sensing errors. Then, we suggest to analyze the performances of the well known Upper Confidence Bound algorithm UCB_1 within this framework, and show that we still can obtain an order optimal channel selection behavior. Finally we compare these results to those obtained in the case of perfect sensing. Simulation results are provided to support the suggested approach.

1 Introduction

1.1 Accès Opportuniste au Spectre

L'accès Opportuniste au Spectre est un concept prometteur, suggéré par la communauté radio, pour exploiter au mieux les opportunités spectrales aujourd'hui disponibles. En effet, durant ce dernier siècle, l'allocation statique des bandes de fréquence aux applications et services sans fil au nombre sans cesse croissant, a mené à une pénurie de la ressource spectrale. Néanmoins, de nombreuses mesures effectuées aux Etats-Unis, d'abord, corroborées ensuite par des études similaires dans le reste du monde, montrent une sous utilisation chronique du spectre [1]. Ces mesures montrent par la même occasion des opportunités de communication substantielles à exploiter.

Le concept général de l'accès opportuniste au spectre définit deux classes d'utilisateurs : les utilisateurs primaires (UP) et les utilisateurs secondaires (US). Les UPs ont accès aux ressources spectrales dédiées à leurs services. Ils sont donc prioritaires sur ces bandes de fréquence. Les USs, par opposition aux UPs, dénotent un groupe d'utilisateurs désireux d'exploiter les opportunités de communication laissées vacantes, à un certain moment dans une certaine zone géographique, par les UPs.

Il est généralement admis que les USs n'ont pas (ou peu) d'information *a priori* sur l'occupation des bandes primaires. De plus, les interférences occasionnées par les USs doivent rester sous un certain seuil toléré par les UPs. Afin de répondre à ces exigences, le concept de la radio intelligente (Cognitive Radio en anglais) a été suggéré [1,2] munissant ainsi les équipements secondaires de capacités cognitives élémentaires, à savoir : observation de l'environnement à travers des capteurs dédiés, analyse des informations collectées, et enfin adaptation du comportement de l'équipement aux fluctuations de l'environnement et aux attentes de l'utilisateur.

Il reste néanmoins de nombreux défis à surmonter afin d'exploiter de manière efficace les opportunités présentes dans le spectre. D'une part, la conception de détecteurs précis et fiables, et d'autre part, l'analyse de mécanismes d'apprentissage et de prise de décision performants. Proposer de tels algorithmes est depuis quelques années au centre de nombreuses recherches [3] [4].

1.2 Le Paradigme des Bandits Manchots

Le paradigme des bandits manchots (Multi-Armed Bandit en anglais) a été récemment le centre d'une attention particulière

de la part de la communauté radio. Brièvement, ce paradigme modélise l'agent de prise de décision par un joueur dans un casino. Ce dernier cherche à maximiser les gains cumulés obtenus en tirant le bras de différentes machines à sous (Multi-Armed Bandit, en anglais). Ces dernières représentent, en revanche, les ressources à exploiter par l'agent intelligent. Si ce joueur avait une information complète sur les gains moyens de chaque machine à sous, une stratégie optimale serait de jouer en permanence la machine avec le gain moyen le plus élevé. Néanmoins, dans la mesure où le joueur ne dispose d'aucune information sur ce qu'il pourrait gagner en jouant telle ou telle machine, il n'a d'autre choix que de tester les différentes machines afin d'estimer leur gain moyen. La recherche d'un équilibre entre le temps passé à tester les différentes machines afin d'estimer leurs performances respectives, et le temps consacré à la machine qui semble être optimale est ce qui est habituellement appelé *dilemme Exploitation-Exploration.* Si nous imaginons que ces machines à sous représentent des bandes spectrales auxquelles l'utilisateur secondaire cherche à accéder, ce problème de décision et d'apprentissage en radio intelligente s'apparente à un problème des bandits manchots.

Ainsi, plusieurs algorithmes ont été empruntés du domaine de l'apprentissage machine [5–7] et suggérés pour répondre à la problématique de l'accès opportuniste au spectre [8–10]. Ces algorithmes, néanmoins suppose une observation sans erreur de l'état des bandes de fréquence. Dans ce contexte, l'utilisateur secondaire peut effectivement maximiser ses gains cumulés sans interférer avec les utilisateurs primaires.

L'objectif de ce papier est d'introduire un modèle plus réaliste. Ainsi, nous considérons un modèle des machines à sous dans le quel des erreurs d'observations sont possibles. Le modèle considéré est détaillé dans la section 2. La section 3 introduit l'algorithme UCB_1 et annonce ses performances. Ces résultats sont illustrés à l'aide de simulations. Enfin la section 4 conclut ce papier.

2 Modélisation de la problématique

2.1 Le réseau primaire

Dans le cadre de l'Accès Opportuniste au Spectre [1,4], nous considérons le cas d'un utilisateur dit "secondaire" intéressé par une bande de fréquence dédiée à un réseau prioritaire dit "réseau primaire". La bande de fréquence d'intérêt est supposée divisée en K sous-bandes indépendantes mais non- identiques.

Soit k l'indice du $k^{\text{ème}}$ canal le plus disponible en probabilité. A chaque fois qu'un canal est sondé, il est observé dans l'un des deux états suivants :{libre, occupé}. Dans le reste du papier, nous associons la valeur numérique 0 à un canal occupé, et 1 autrement. L'occupation temporelle d'une sous-bande k est supposée suivre une distribution de Bernoulli inconnue θ_k. De plus les distributions $\Theta = \{\theta_1, \theta_2, \cdots, \theta_K\}$ sont supposées stationnaires.

Ce papier s'adresse au cas particulier d'un réseau primaire synchrone où le temps $t = 0, 1, 2, \cdots$, est supposé divisé en paquets de taille fixée. Notons $\mathbf{S}_t$ l'état des canaux à l'itération $t : \mathbf{S}_t = \{S_{1,t}, \cdots, S_{K,t}\} \in \{0,1\}^K$. Pour tout $t \in \mathbb{N}$, la valeur numérique $S_{k,t}$ est supposée être la réalisation aléatoire de la distribution θ_k. De plus, les réalisations $\{S_{k,t}\}_{t\in\mathbb{N}}$ tirées de la distribution θ_k sont supposées indépendantes et identiquement distribuées. La disponibilité moyenne d'un canal est caractérisée par sa probabilité d'être libre. Ainsi, nous définissons la disponibilité μ_k du canal k telle que pour tout $t : \mu_k = \mathbb{P}(S_{k,t} = 1)$, avec $\mu_1 > \mu_2 \geq \cdots \geq \mu_k \geq \cdots \geq \mu_K$ sans perte de généralité.

2.2 L'utilisateur secondaire

Nous décrivons dans ce paragraphe le moteur de prise de décision ainsi que les caractéristiques du détecteur de signaux associés à l'utilisateur secondaire.

Nous dénommons "Agent Intelligent" (AI) le moteur de prise de décision de l'équipement de radio intelligente. L'AI peut être vu comme le centre névralgique de l'équipement. A chaque paquet t, il doit choisir un canal à observer. Pour cela, les décisions de l'AI reposent sur les informations passées collectées au fur et à mesure de ses interactions avec l'environnement. Soit i_t le vecteur "information" disponible à l'instant t. Nous supposons que l'AI ne peut observer qu'un canal à la fois à chaque itération t. Ainsi, le choix d'un canal à observer peut être associé à une action $a_t \in \mathcal{A}$ où l'ensemble $\mathcal{A} = \{1, 2, \cdots, K\}$ fait référence à l'ensemble de canaux considéré par l'utilisateur secondaire. Par conséquent l'AI peut être considéré en tant que fonction de décision π qui associe pour tout t, une action a_t à l'information $i_t : a_t = \pi(i_t)$

Soit $X_t \in \{0,1\}$ la réalisation aléatoire calculée à l'issue de l'étape d'observation à l'instant t du canal sélectionné a_t. Dans le cas d'une détection parfaite (sans erreur d'observation), nous aurions : $X_t = S_{a_t,t}$. Dans le contexte considéré, cependant, la valeur de X_t est fonction des caractéristiques opérationnelles du récepteur (COR). Les COR définissent la précision et la fiabilité d'un détecteur via la mesure de deux types d'erreurs : d'une part la détection d'un utilisateur primaire sur la bande sondée alors que la bande est en réalité libre. Cette erreur est communément appelée "fausse alarme". D'autre part, la "détection manquée" revient à considérer une bande libre alors qu'elle est occupée par des utilisateurs primaires à l'instant t. Notons ϵ et δ, respectivement, les probabilités de fausse alarme et de détection manquée caractérisant l'équipement de radio intelligente considéré lors de cette étude :

$$\begin{cases} \epsilon = \mathbb{P}_{fa} = \mathbb{P}(X_t = 0 | S_{a_t,t} = 1) \\ \delta = \mathbb{P}_{md} = \mathbb{P}(X_t = 1 | S_{a_t,t} = 0) \end{cases}$$

Finalement, le résultat de l'étape d'observation peut être associé à la sortie d'une fonction aléatoire $\pi_s(\epsilon, \delta, S_{a_t,t})$ tel que : $X_t = \pi_s(\epsilon, \delta, S_{a_t,t})$. Cependant, nous ne nous intéressons pas aux formes possibles de fonctions de détection et renvoyons le lecteur intéressé à la référence suivante : [3].

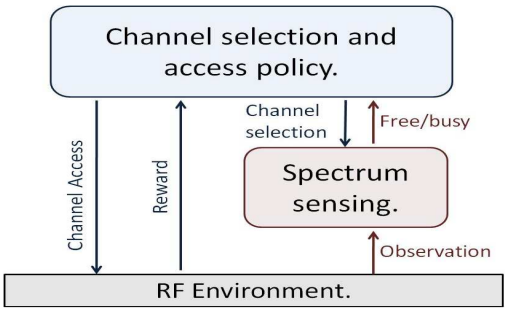

FIGURE 1 – Représentation de l'interaction d'un agent intelligent avec son environnement RF.

2.3 Stratégie d'accès au canal

En fonction du résultat de l'étape d'observation $X_t \in \{0, 1\}$, l'AI peut choisir d'accéder ou pas au canal sélectionné. Soit $\pi_a(X_t) \in \{0, 1\}$ la décision d'accès au canal, telle que, d'une part, 0 désigne un refus d'accès et, d'autre part, 1 désigne une autorisation d'accès au canal. Pour des raisons de simplicité, la stratégie choisie dans ce papier se résume à "accéder le canal s'il est observé libre". En d'autres termes : $\pi_a(X_t) = X_t$. Nous supposons dans ces travaux que le détecteur est modélisé de manière à assurer un niveau d'interférence, avec l'utilisateur primaire, inférieur à un certain seuil fixé par les réglementations. Néanmoins, nous ne supposons pas nécessairement connus les paramètres $\{\epsilon, \delta\}$. De plus nous supposons qu'il existe un mécanisme permettant à l'utilisateur secondaire d'être informé en cas d'interférence avec l'utilisateur primaire. Dans ce cas, la transmission de l'utilisateur secondaire est assimilée à un échec. Enfin, nous supposons qu'un paquet de taille D_t est envoyé à chaque tentative de transmission (i.e., accès au canal). Ainsi, à la fin de chaque paquet t, l'AI calcule une valeur numérique r_t, habituellement appelée *gain* dans la communauté de l'apprentissage machine, qui quantifie les performances instantanées du moteur de prise de décision de l'utilisateur secondaire.

L'interaction de l'utilisateur secondaire avec son environnement est résumée et illustrée dans la figure 1.

3 Algorithmes et performances

3.1 Algorithme de selection du canal

Nous analysons dans ce papier l'impact des erreurs d'observation sur les performances de l'algorithme UCB_1. Ce dernier avait été précédemment suggéré pour sa simplicité et la garantie de ses propriétés mathématiques de convergence [6] [7] [9]. Brièvement, cet algorithme repose sur l'affectation d'un indice de qualité à chaque canal en fonction des précédentes observations et tentatives de transmissions. La forme de l'index considéré dans ce papier est la suivante :

$$B_{k,t,T_k(t)} = \overline{X}_{k,T_k(t)} + A_{k,t,T_k(t)}$$

Dans cette expression, d'une part $\overline{X}_{k,T_k(t)} = \frac{\sum_{m=0}^{t-1} r_m \cdot \mathbf{1}_{\{a_m=k\}}}{T_k(t)}$ représente la moyenne empirique des gains obtenus à partir du canal k, après $T_k(t)$ tentatives au bout de t itérations, d'autre part, $A_{k,t,T_k(t)}$ représente un biais ajouté afin d'assurer la convergence de l'algorithme vers le canal optimal. Dans le cadre de cette étude, le biais considéré est le suivant :

$$A_{k,t,T_k(t)} = \sqrt{\frac{\alpha . \ln(t)}{T_k(t)}}$$

où α est un paramètre (réel positif) d'apprentissage. Finalement, l'algorithme π, de sélection des canaux, choisit à chaque itération t le canal avec la plus grande valeur associée à l'indice $B_{k,t,T_k(t)}$, tel que :

$$a_t = \pi(i_t) = \arg\max_k (B_{k,t,T_k(t)})$$

Une version détaillée de l'implémentation de cet algorithme avait déjà été proposée dans une précédente étude [9].

3.2 Performances

En vue du modèle introduit précédemment, nous considérons un gain de la forme suivante :

$$r_t \triangleq D_t S_{a_t,t} \pi_a(X_t)$$

Néanmoins pour des raisons de simplicité, nous considérons qu'à chaque tentative de transmission, l'utilisateur secondaire transmet $D_t = 1$ bit. Par conséquent, en tenant compte de la forme de la stratégie d'accès, et des simplifications introduites, le gain r_t peut s'exprimer de la manière suivante :

$$r_t = S_{a_t,t} X_t$$

On montre que l'espérance du temps $\mathbb{E}[T_k(t)]$ passé par l'algorithme à sélectionné un canal sous optimal (i.e., $k \in \{2, \cdots, K\}$) est borné par une fonction logarithmique du nombre d'itérations (*slot* en anglais) tel que pour $\alpha > 1$ et $\Delta_k = \mu_1 - \mu_k$:

$$\mathbb{E}[T_k(t)] \leq \frac{4\alpha \ln(t)}{((1-\epsilon)\Delta_k)^2} \quad (1)$$

La preuve de ce résultat est une extension des travaux réalisés dans les papiers [6, 7]. Les techniques utilisées pour mener la preuve sont aussi similaires. Pour des raisons d'espace, nous ne prouverons pas ce résultat dans ce papier. Il est néanmoins possible d'y accéder dans le papier [11].

Ce résultat montre que malgré les erreurs d'observation, il est toujours possible de converger rapidement vers le canal optimal, néanmoins, sans surprise, nous observons une dégradation de la vitesse de convergence qui est directement liée aux performances du détecteur.

Ce résultat a été, de plus, confirmé en simulation, tel que illustré par la figure 2. Cette figure suppose la disponibilité de dix canaux avec des probabilités de disponibilités μ_k, $k \in \{1, 2, \cdots, K\}$ respectivement égaux à $[0.9, 0.8, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1]$. Ces résultats sont moyennés sur 100 réalisations de l'expérience. On voit notamment que l'algorithme a une

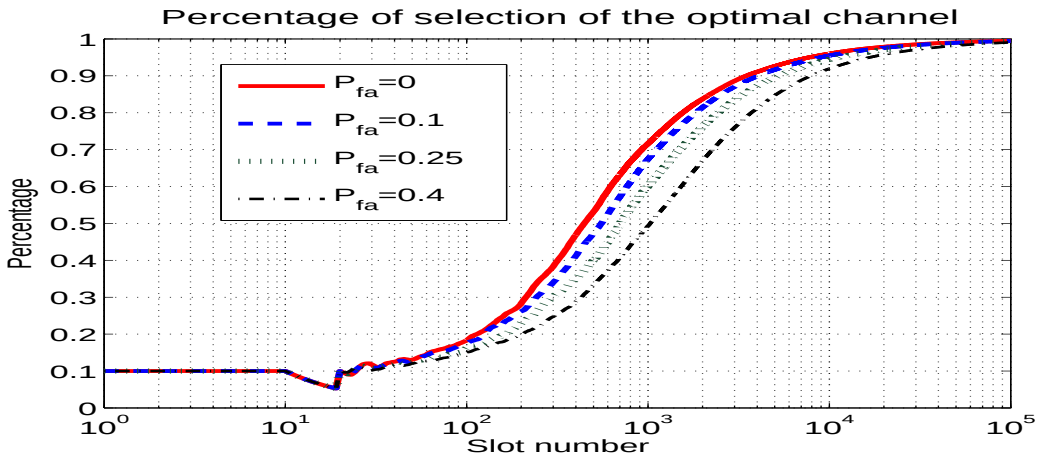

FIGURE 2 – Pourcentage du temps passé à sélectionner le canal optimal à l'aide de l'algorithme UCB_1 en fonction des erreurs d'observations.

première phase d'exploration pendant laquelle aucun canal n'est favorisé. Ensuite, nous observons que la courbe croit rapidement mettant en avant la capacité de l'algorithme à déterminer le canal optimal quelque soit les erreurs d'observations. Néanmoins, plus le détecteur est précis, meilleurs est la phase d'exploitation de l'algorithme illustrée par le temps de convergence de l'AI.

4 Conclusion

Nous avons introduit, dans ce papier, un modèle d'accès opportuniste au spectre sous la forme d'un problème de machines à sous. Ce dernier a été complété par un modèle des canaux prenant en compte les erreurs d'observation. En d'autres termes, des erreurs d'observation de l'état des canaux peuvent avoir lieu aléatoirement à chaque mesure. Ensuite, les performances de l'algorithme UCB_1 ont été analysées avant d'être validées par des simulations. Ainsi, ce papier montre que malgré les erreurs d'observations qui peuvent avoir lieu lors de la phase de détection des utilisateurs primaires, l'algorithme UCB_1 reste capable d'apprendre et de converger vers le canal optimal. Ainsi, en fonction de la précision de la détection, la convergence de l'algorithme peut être plus ou moins rapide.

Bien que ces résultats soient encourageants, de nombreux questionnements résident quant à leur généralisation. En effet dans le cas de réseaux d'utilisateurs secondaires, il est impératif de s'assurer que ces derniers n'interfèrent pas entre eux afin de ne pas ruiner la phase d'apprentissage. Ce dernier point est actuellement en cours d'étude.

Références

[1] Federal Communications Commission. Spectrum policy task force report. November 2002.

[2] J. Mitola and G.Q. Maguire. Cognitive radio : making software radios more personal. *Personal Communications, IEEE*, 6 :13–18, August 1999.

[3] T. Yucek and H. Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *In IEEE Communications Surveys and Tutorials*, 11, no.1, 2009.

[4] Q. Zhao and B. M. Sadler. A survey of dynamic spectrum access : signal processing, networking, and regulatory policy. In *in IEEE Signal Processing Magazine*, pages 79–89, 2007.

[5] R. Agrawal. Sample mean based index policies with O(log(n)) regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27 :1054–1078, 1995.

[6] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite time analysis of multi-armed bandit problems. *Machine learning*, 47(2/3) :235–256, 2002.

[7] J.-Y. Audibert, R. Munos, and C. Szepesvári. Tuning bandit algorithms in stochastic environments. In *Proceedings of the 18th international conference on Algorithmic Learning Theory*, 2007.

[8] L. Lai, H.E. Gamal, H.J. Jiang, and V. Poor. Cognitive medium access : Exploration, exploitation and competition. *[Online]. Available : http ://arxiv.org/abs/0710.1385*.

[9] W. Jouini, D. Ernst, C. Moy, and J. Palicot. Upper confidence bound based decision making strategies and dynamic spectrum access. *Proceedings of the 2010 IEEE International Conference on Communications (ICC)*, May 2010.

[10] K. Liu and Q. Zhao. Distributed learning in cognitive radio networks : Multi-armed bandit with distributed multiple players. 2010.

[11] W. Jouini, C. Moy, and J. Palicot. Upper confidence bound algorithm for opportunistic spectrum access with sensing errors. *6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, Osaka, Japan*, June 2011.