

Accès Opportuniste dans les systèmes de communication multi-canaux : Une perspective du problème de bandit-manchot

Introduction

Dans les dernières décennies, les réseaux sans fil ont permis le déploiement d'un vaste ensemble de systèmes de communication avancés, tels que la communication mobile, un accès Internet mobile et sans fil des données des capteurs de récolte. La densité de dispositifs de l'accès sans fil et la demande pour les communications sans fil devraient augmenter de façon spectaculaire, ce qui nous amène dans l'omniprésence de l'informatique et de l'environnement de communication, nommés par l'Union européenne "Internet des Objets".

La radio cognitive, d'abord envisagée par Mitola puis étudiée par le programme de DARPA XG, est la technologie clé habilitante pour les générations futures de systèmes sans fil qui répondent aux défis critiques en matière d'efficacité du spectre, la gestion des interférences, et de la coexistence de réseaux hétérogènes. Le concept de base dans les réseaux de radio cognitive est l'accès au spectre opportuniste (OSA), dont l'objectif est de résoudre le déséquilibre entre la rareté du spectre et le spectre de sous-utilisation. L'idée de base de l'OSA est de permettre aux utilisateurs secondaires pour rechercher, identifier et exploiter les possibilités du spectre instantanées tout en limitant l'interférence perçue par les utilisateurs primaires. Construit sur une structure hiérarchique avec un accès aux utilisateurs primaires et secondaires, OSA résout l'inefficacité du modèle actuel de commandement et de contrôle de la réglementation du spectre, tout en maintenant la compatibilité avec les systèmes sans fil existants.

Bien que conceptuellement simple, OSA dans les réseaux de radio cognitive présente des défis nouveaux qui ne sont pas présentés dans les réseaux classiques, tels que la détection sur une large bande de fréquence, l'identification de la présence des utilisateurs primaires, la détermination de la nature des possibilités, et la coordination de l'utilisation de ces possibilités avec d'autres noeuds sans interférer avec les principaux utilisateurs.

Dans cette thèse, nous nous concentrons sur l'accès au spectre opportuniste où un utilisateur secondaire tente d'identifier des opportunités dans l'espace blanc spectral lorsque les utilisateurs primaires ne sont pas actifs. L'utilisateur secondaire étant depuis premier à effectuer la détection de spectre, qui est généralement faite comme une détection de l'énergie, à savoir, pour déterminer si l'énergie observée par l'utilisateur secondaire peut être expliquée par l'énergie du bruit seul ou si elle indique la présence d'un signal primaire. En raison de la contrainte matérielle et le coût de détection, le délai et l'énergie, il est impossible pour l'utilisateur secondaire de détecter tous les canaux candidats. Par conséquent, dans les circonstances partiellement observables, l'utilisateur secondaire doit décider quels canaux devraient être détectés dans l'intervalle de temps et si le canal détecté devrait être jugé inactif sur la base des imparfaits de détection, et choisir en outre un ensemble de canaux inactifs et transmet des informations sur eux. Nous explorons comment l'utilisateur secondaire devrait exploiter les observations passées et la connaissance des propriétés stochastiques de ces canaux afin de maximiser son taux de transmission par commutation de canaux de façon opportuniste.

La conception de la politique optimale de détection peut être formulée comme un processus de décision de Markov partiellement observable (POMDP), ou d'un problème de repos de plusieurs bandit manchot (RMAB), dont l'application est bien au-delà du domaine des systèmes de radio cognitive. Malheureusement, l'obtention de la politique optimale pour un POMDP général ou pour un processus agité bandit multi-armée est souvent intraitable en raison de la complexité de calcul exponentielle qui est prouvé être PSPACE-dur par Papadimitriou et Tsitsiklis en 1999.

En raison de son caractère générique, le problème RMAB revêt une importance fondamentale dans la théorie de décision stochastique et trouve son application dans de nombreux problèmes d'ingénierie tels que l'accès au canal sans fil, le brouillage de communication et de suivi d'objet. Cependant, malgré les importants efforts de recherche dans le domaine, le problème RMAB dans sa forme générique reste encore ouverte. Jusqu'à aujourd'hui, très peu de résultat est rapporté sur la structure de la politique optimale.

Par conséquent, les efforts de recherche au s'attaquer problème RAMB découlant de diverses applications, des systèmes de communication en particulier et des réseaux des systèmes dynamiques tombent généralement dans les trois catégories suivantes. La première consiste à rechercher des conditions suffisantes pour les politiques simples et robustes (par exemple, la politique à vue myopique, politique gloutonne) en vertu

de laquelle l'optimalité de ces politiques est garanti. Le second est de construire des politiques particulières dont les performances de l'optimal sont délimitées. La troisième est, en suivant la ligne de recherche de Whittle, consiste à calculer l'indice de Whittle et d'en tirer des politiques fondées sur l'indice Whittle.

Dans cette thèse, nous fournissons une analyse générique sur le problème d'accès au canal opportuniste en lançant le problème dans le problème de repos multi-bandit manchot (RMAB). Du point de vue technique, le projet RMAB est considérablement complexe car il implique non seulement l'ordre de détection, mais aussi le nombre de détection du canal, où la forme signifie quels canaux qui devraient être détectés et accessibles à chaque emplacement, c'est à dire, de choisir le canal dans l'ordre descendant la probabilité de disponibilité du canal ou dans un ordre aléatoire, tandis que le second exige que l'utilisateur secondaire de décider combien de canaux devraient être détectés au niveau de chaque emplacement.

Par conséquent, pour la facilité de l'analyse, nous analysons d'abord la détection du problème supposant que le nombre de détection du canal est fixé dans la période de décision. En particulier, nous nous concentrons sur l'ordre intuitif de détection, également appelé la politique myopique, uniquement en maximisant la récompense immédiate, tout en ignorant l'impact d'une récompense future, et d'obtenir des conditions suffisantes pour garantir l'optimalité de la politique myopique dans la première partie de la thèse, et dans la partie restante, dans l'hypothèse de l'ordre intuitif de détection, nous abordons le nombre optimal de détection de la voie et nous proposons une politique d'anticipation heuristique pour améliorer l'efficacité de détection.

Optimalité de la politique de détection myope dans OSA : une analyse de motivation

Selon la partie précédente, l'idée de base de l'accès du spectre opportuniste (OSA) est d'exploiter la disponibilité du spectre instantanée en permettant à un utilisateur pour accéder aux chaînes temporairement inutilisées de façon opportuniste. Dans ce contexte, une politique bien conçue de canal d'accès est essentiel pour atteindre l'utilisation efficace du spectre.

Plus précisément, considérons un scénario générique où il y a N canaux spectraux à fente. Chaque canal évolue comme un indépendant et identiquement distribué (iid), deux Etats en temps discret de la chaîne de

Markov. Les deux états pour chaque canal, mauvais (état 0) et bon (état 1), indique si le canal est libre pour l'utilisateur de transmettre son paquet sur ce canal à un slot donné. Les probabilités de transition d'état sont données par p_{ij} , $i, j = 0, 1$. L'utilisateur cherche une politique de détection qui exploite de manière opportuniste les canaux inutilisés temporairement à transmettre ses paquets. À cette fin, dans chaque slot, l'utilisateur sélectionne un sous-ensemble de canaux pour capter sur la base de ses observations antérieures et d'obtenir une unité comme récompense si au moins l'un des canaux détectée est dans le bon état, indiquant que l'utilisateur peut effectivement envoyer un paquets en utilisant le canal inactif (ou l'un des canaux inactifs) dans le slot courant. L'objectif de l'utilisateur est de trouver la politique optimale de détection qui maximise la récompense qu'il peut obtenir sur un horizon fini ou infini.

La conception de la politique optimale de détection peut être formulée comme un processus de décision de Markov partiellement observable (POMDP), ou un problème bandit agité à plusieurs bras (RMAB), dont l'application est bien au-delà du domaine des fonctions cognitives les systèmes radio. Malheureusement, l'obtention de la meilleure politique pour un processus général POMDP ou agité bandit multi-armée est souvent intraitable en raison de la complexité de calcul exponentiel. Par conséquent, une alternative naturelle est de chercher simples politiques myopes pour l'utilisateur. Dans cette ligne de recherche, une stratégie myope de détection est développée pour le cas où l'utilisateur est limité à détecter un seul canal à chaque slot. La politique à courte vue de détection myopic dans ce cas vas être optimale lorsque $p_{11} > p_{01}$.

Dans cette thèse, nous avons naturellement d'étendre le projet de politique myopique pour le cas générique où l'utilisateur peut ressentir plus d'un canal dans chaque slot et obtient une unité de récompense si au moins l'un des canaux détectés est inactif. L'analyse théorique montre que la politique généralisée myope n'est optimale que pour un petit sous-ensemble de cas où l'utilisateur sens deux canaux dans chaque fente. Dans les cas génériques, nous donnons des contre-exemples pour montrer que la politique à courte vue, en dépit de sa structure simple, n'est pas optimale. Il est instructif de comparer nos résultats obtenus dans ce chapitre avec une autre extension parallèle sur le même problème. Les auteurs montrent que, lorsque $p_{11} > p_{01}$ détient, la politique à courte vue de détection est optimale, même pour le cas où les sens des utilisateurs plus d'un canal dans chaque fente. Toutefois, ce résultat semble être en contradiction avec notre conclusion. En fait, cette contradiction est due au fait que l'objectif de l'utilisateur est de trouver le plus grand nombre de bons canaux que possible afin que l'utilisateur peut transmettre sur tous les canaux inactifs. En revanche, nos résultats sont axés

sur le scénario où l'utilisateur ne peut transmettre sur un canal libre, même si plusieurs canaux inactifs peuvent exister dans le slot courant. Dans un autre mot, l'utilisateur vise à maximiser la probabilité de trouver au moins un bon canal. Il est instructif de remarquer que la nuance sur le modèle (plus spécifiquement sur la fonction d'utilité) conduit en effet à des résultats totalement contraires, ce qui indique que les efforts de recherche supplémentaires sont nécessaires pour comprendre les caractéristiques intrinsèques de la politique à courte vue et la politique optimale, ce qui motive notre travail présenté dans cette thèse.

Une analyse axiomatique sur l'optimalité de la politique de détection myope

Motivé par l'analyse ci-dessus, nous étudions ensuite le naturel qui suit tout en question fondamentalement importante: dans quelles conditions est la politique myope garantie d'être optimale? Nous répondons à la question ci-dessus posée par l'exécution d'une étude axiomatique. Plus précisément, nous développons trois axiomes---symétrie, décomposabilité et monotonie---caractériser une famille de fonctions génériques et important que nous appelons en tant que fonctions régulières. Nous établissons ensuite l'optimalité de la politique à courte vue lorsque la fonction récompense peut être exprimée comme une fonction régulière et le facteur d'actualisation est limitée par un seuil fermé forme déterminée par la fonction récompense. Nous avons ensuite illustrer la façon dont les résultats sont dérivés, de nature générique, sont appliqués à analyser une classe de problèmes qui se posent dans l'accès RMAB multi-canal opportuniste dans deux scénarios.

Le premier modèle dans les réseaux radio cognitifs implique le problème afin de détection où l'utilisateur sélectionne secondaire k ($1 < k < n$) de n canaux afin de maximiser la probabilité pour trouver un canal libre. Il est évident que la fonction récompense immédiate est l'ordre de 1 combinaison non linéaire de la probabilité de disponibilité des canaux sélectionnés. Le résultat démontre que la politique gloutonne n'est pas optimale, généralement sous la récompense moyenne prévu.

Le second modèle est que l'utilisateur choisit k ($0 < k < n$) de voies d'accès et de recevoir une récompense sur la voie en bon état. La fonction de récompense immédiate est la combinaison linéaire de la probabilité de disponibilité de ces canaux sélectionnés. Notre résultat obtenu est cohérent avec ceux où la politique à courte vue de choisir n'importe quel nombre de canaux est optimale.

Afin d'étudier plus avant l'optimalité de la politique à courte vue dans le cas de modèle de canal asymétrique, nous étendons la fonction régulière à g-régulière fonction en étendant la décomposabilité à g-décomposabilité. Axiome 3 sur les Etats g-décomposabilité que $f(w_i)$ peut toujours être décomposé en deux termes, en introduisant la fonction g et le remplacement w_i par 0 et 1, respectivement. Il est perspicace de noter Axiom de g-décomposabilité étend considérablement Axiom de la décomposabilité en couvrant un éventail beaucoup plus large de fonctions utilitaires qui ne peuvent pas être couverts par formel, en particulier la fonction logarithmique qui sont largement utilisés dans des problèmes d'ingénierie. En fixant $g(w_i) = w_i$ et $c = 1, 3$ Axiom dégénère à l'axiome de décomposabilité. Puis, par la voie de distinction que les canaux détectés et non détectés canaux à chaque emplacement, nous avons quantifié, sous forme fermée, le compromis de l'exploration vs exploitation dans le processus de décision séquentiel, puis dérivée de la condition suffisante pour l'optimalité de la politique à courte vue .

En comparaison avec la littérature existante aborder l'optimalité de la politique à courte vue du problème RMAB, la contribution de notre analyse est résumée comme suit :

1. Analyse générique : Contrairement à la ligne de recherche visant à démontrer l'optimalité ou non-optimalité de la politique à courte vue dans les scénarios d'application indiquées, nous faisons un effort plus générique sur les conditions suffisantes assurant l'optimalité de la politique à courte vue.
2. Canaux homogènes et hétérogènes : Nous étudions les deux scénarios homogènes et hétérogènes où les canaux sont caractérisés par des chaînes de Markov homogènes et hétérogènes, respectivement.
3. Sentant l'erreur : La grande majorité des études dans le domaine supposent l'observation parfaite des Etats canal. Cependant, des erreurs de détection ou d'observation sont inévitables dans le scénario pratique (par exemple, en raison des limites de bruit et le système), en particulier dans les systèmes de communication sans fil. Notre travail reflète l'erreur de détection et étudie l'optimalité de la politique à courte vue en vertu de la détection imparfaits.

Du point de vue méthodologique, nous adoptons une approche axiomatique de rationaliser l'analyse. D'une part, une telle approche axiomatique fournit une vue hiérarchique du problème abordé et conduit à l'analyse plus claire et plus synthétique. D'autre part, l'approche axiomatique permet également de réduire la complexité de la résolution du problème RMAB et illustre certaines des implications importantes en génie derrière la politique à

courte vue.

Au-delà de détection myope : une heuristique v-étape politique Lookahead

D'autre part, nous avons encore enquêter sur un problème plus difficile, où l'utilisateur doit décider du nombre de canaux à sens afin de maximiser son utilité.

Dans l'hypothèse où un utilisateur cognitive ne peut pas accéder à tous les canaux disponibles simultanément, la tâche principale du protocole d'accès au support est de choisir les canaux que l'utilisateur cognitive devrait tenter d'utiliser dans des créneaux différents, afin de tirer pleinement (ou au maximum) utiliser les possibilités spectrales. Ce processus de décision peut être améliorée prenant en compte toutes les données statistiques sur le trafic primaire. Par exemple, avec une capacité de détection des canaux multiples dans une fente, un sens des utilisateurs secondaires de ces canaux de façon séquentielle selon l'ordre de détection certaine, c'est à dire, l'ordre décroissant de probabilité de disponibilité du canal, afin de recueillir des informations jusqu'à ce qu'il s'arrête sur un canal sur la base de certains des critères tels que l'obtention de succès d'un canal disponible, puis transmet ses informations pendant le reste de cet emplacement.

Il ya un ensemble de littératures connexes portant sur l'ordre de détection et le problème d'arrêt optimal où un utilisateur secondaire peut séquentiellement et canaux sensoriels en continu sélectionnés jusqu'à ce qu'un canal est identifié pour le repos. Le canal de détection optimale stratégie, y compris l'ordre de détection et de la règle d'arrêt a été calculée pour un seul utilisateur le cas avec une hypothèse de rappel et deviner où l'ancien permet à l'utilisateur secondaire d'accéder à un canal précédemment détectée alors que les permis de ce dernier l'utilisateur secondaire pour accéder à un canal qui n'a pas encore été détecté. Il a été démontré que l'obtention du canal de détection optimale la stratégie est de calcul prohibitifs, puis polynomial complexité des algorithmes avec le paramètre v ont reçu des récompenses au plus v inférieure à celle de la stratégie optimale. La détection optimale d'ordre problème a été étudié pour un utilisateur unique cas où ni rappelant ni deviner est accueilli avec un simple ordre de détection pour certains scénarios particuliers. Un coordonnateur centralisé a été utilisé pour découvrir les possibilités du spectre pour un réseau radio cognitive et de l'ordre de détection canal optimal a été dérivées pour les canaux avec des capacités homogènes, et il a été montré que le problème

de l'ordre optimal de détection pour les canaux avec des capacités hétérogènes est NP-difficile.

En général, l'objectif du canal de détection est de trouver un canal disponible le plus rapide que possible. Nous supposons que l'utilisateur doit détecter les canaux selon l'ordre décroissant de probabilité de disponibilité du canal, dénommé afin de détection intuitive, où il est prouvé à être optimale lorsque la modulation adaptative n'est pas utilisée. Cependant, de détection dans l'ordre intuitif est un type de comportement myope si l'utilisateur arrête de détection une fois un canal disponible est obtenu pour son compte que le coût de détection dans la fente en cours, mais en ignorant que, dans la case suivante. Pour surmonter la lacune du comportement myope, l'utilisateur peut ressentir plus de canaux pour parvenir à un compromis fondamental entre l'exploitation et l'exploration. Surtout, l'exploitation se réfère à la transmission d'informations sur ce canal senti l'oisiveté, sans détection des canaux supplémentaires alors que l'exploration est le processus d'apprendre le comportement statistique du trafic primaire par la détection des chaînes supplémentaires, même après l'obtention d'un canal disponible, ce qui profite la décision dans les fentes suivantes. Par conséquent, lorsque l'utilisateur arrête de détection des canaux supplémentaires après avoir trouvé un canal disponible est le point clé dans la phrase de détection d'une fente.

Dans cette thèse, nous prenons une première étape vers l'élaboration d'un cadre pour l'analyse de détection de problème d'arrêt dans des environnements incertains. Notre approche permet la construction de stratégies qui établissent un équilibre entre l'exploitation et l'exploration sous certaines hypothèses de modélisation. Compte tenu de la politique à courte vue de détection, le problème d'arrêt optimal est coulé dans un problème de bandit à plusieurs bras agités dont il est prouvé à la PSPACE-dur. Ensuite, nous développons une méthode heuristique v-étape politique d'anticipation qui se compose de détection des canaux d'une manière myope et l'arrêt de la détection lors de l'utilité attendue agrégées à partir des t sous actuels à la fente $t+v$ commence à diminuer. Dans la politique développée, le paramètre v permet de parvenir à un compromis souhaité entre l'efficacité sociale et de la complexité de calcul. Il est instructif de noter que le projet de v-étape politique d'anticipation peut être décomposée en deux étapes :

1. Exploitation: l'utilisateur exploite les informations actuellement disponibles d'une manière gourmande de manière à trouver un bon canal ;
2. Exploration: une fois un bon canal sécurisé, l'utilisateur procède à explorer l'espace d'état système de gain à long terme.

La deuxième étape (l'exploration) peut être absent si tous les meilleurs canaux de M sont détectés mauvaise ou si l'exploration n'augmente pas le gain à long terme (c.-à-d. la condition dans l'algorithme ne tient pas même une seule fois).

Nous démontrons les avantages de la stratégie proposée par l'intermédiaire d'expériences numériques sur plusieurs paramètres typiques. En analysant deux scénarios typiques --- canaux symétriquement positivement corrélés et les canaux asymétriques, nous obtenons que le régime prévu se comporte bien dans les cas de voies positivement corrélés. Par conséquent, la règle d'arrêt peuvent être résumées comme suit: l'utilisateur secondaire doivent utiliser le système d'anticipation pour améliorer l'efficacité de détection dans le cas de canaux positivement corrélés tout en abandonner le régime d'anticipation dans les autres cas, qui est, de transmettre des informations une fois un canal disponible est détectée.

Conclusion et perspectives

Dans cette section, nous discutons de certaines questions clés ouvertes et esquisser quelques orientations possibles pour des recherches plus poussées.

Accès aux canaux RMAB base avec plusieurs utilisateurs

Dans cette thèse, nous nous concentrons principalement sur le processus décisionnel et compromis différent au sein d'un seul utilisateur. Une orientation de la recherche naturelle est de prendre les résultats obtenus dans la thèse en tant que bloc de construction pour étudier plus avant le scénario de plusieurs utilisateurs accèdent de façon opportuniste un système de communication multi-canal. Ici, le défi de la recherche est de savoir comment coordonner les utilisateurs d'accéder à différents canaux de façon distribuée sans ou avec peu explicite au niveau du réseau de rétroaction. Une façon naturelle de s'attaquer à ce problème consiste à modéliser la situation comme un jeu non coopératif entre les utilisateurs et de voir comment les résultats obtenus dans cette thèse peuvent en outre être adaptée dans le nouveau contexte. Nous commençons maintenant à réaliser des expériences numériques sur le canal de détection et d'accéder à des stratégies développées dans la thèse dans le cadre de plusieurs utilisateurs et d'explorer le problème du point de vue du jeu de signalisation.

Intégrer le coût de commutation

Un autre aspect qui peut limiter la performance du mécanisme d'accès au canal est le coût de commutation de canal. Dans les dispositifs sans fil actuels, le canal de commutation présente un coût en termes de retard, de perte de paquets et les frais généraux du protocole. Par conséquent, une politique canal efficace d'accès doit

éviter de commutation de canal fréquemment, à moins nécessairement. Dans le cadre de RMAB, ce problème peut être mappé sur le problème RMAB générique avec commutation de coût entre les bras. Notre analyse dans le chapitre 6 est une étape primaire vers la prise du canal de commutation coût en compte, mais des travaux plus systématiques sont nécessaires afin de fournir plus en profondeur un aperçu sur ce problème.

Il est important de noter que le générique du MAB avec commutation coût est NP-difficile et il n'existe pas une politique indice optimal. Plus précisément, l'introduction de commutation coût rend non seulement le sous-optimale Gittins politique indice, mais aussi rend la politique optimale de calcul prohibitif. Compte tenu de ces difficultés, nous envisageons d'aborder le problème sous les aspects suivants :

1. Vous recherchez de la politique sous-optimale à la perte de l'efficacité limitée par rapport à la politique optimale ;
2. Developing politique heuristique parvenir à un compromis entre optimalité et de la complexité, que celle présentée dans le chapitre 6 ;
3. Issu politique optimale dans un sous-ensemble de scénarios ou de conception de la politique asymptotiquement optimal.

RMAB avec armes corrélées

Une autre extension pratique est de considérer les canaux corrélés, c'est-à-dire, les chaînes de Markov de différents canaux peuvent être corrélées. Ce problème peut être jeté dans le problème à bras RMAB corrélés. L'introduction de la corrélation entre les bras rend le compromis entre exploration et l'exploitation plus sophistiqué que la détection d'un canal peut non seulement révéler l'état du canal détectée, mais aussi fournir des informations sur d'autres chaînes comme ils ne sont pas tout à fait indépendante. Comment caractériser le compromis dans ce nouveau contexte et la façon de concevoir la politique efficace d'accès au canal sont bien sûr en appuyant sur des sujets de recherche dans ce sens.