Decisional process for ad hoc networks.

Luca Rose

Supélec

STITS ECOLE DOCTORALE

N° d'ordre : 2014-01-TH

THÈSE DE DOCTORAT

Domaine : STIC - Spécialité : Télécommunications

École doctorale "Sciences et Technologies de l'Information, des Télécommunications et des Systèmes"

Présentée par :

Luca ROSE

Sujet:

Processus de décision pour réseaux ad hoc

(Decisional process for ad hoc networks)

Soutenue le 24 janvier 2014 devant les membres du jury:

M. Eitan Altman,	INRIA	Examinateur
M. Mérouane Debbah,	Supélec	Examinateur, Encadrant de Thèse
M. Pierre Duhamel,	CNRS/Supélec	Examinateur, Président du Jury
M. Rida Laraki,	École Polytechnique	Examinateur
M. Christophe J. Le Martret,	Thales Communications & Security	Examinateur, Encadrant de Thèse
M. Marco Luise,	Università di Pisa,	Rapporteur
M. Bruno Tuffin,	INRIA	Rapporteur

Contents

Acknowledgments

I would like to express my deep gratitude to my PhD advisor, Mérouane Debbah, without whom this adventure would not have been possible. I thank him deeply for his excellent guidance, his constant encouragement, and his availability throughout this journey. I appreciate mostly his enthusiasm, passion for research, his hard-working mindset, and immense knowledge. I could not have imagined having a better advisor for my thesis.

My second thought goes to Christophe J. Le Martret, my thesis co-advisor. I thank him sincerely for accepting me as a candidate for this thesis. I also thank him for all his advices, insightful comments, and useful discussions throughout these last three years. Moreover, I thank him for having well supported and encouraged me in speaking in French at the beginning of my thesis. I sincerely thank Marco Luise, Bruno Tuffin, Eitan Altman, and Rida Laraki for accepting to be members of my defense jury, and Pierre Duhamel for accepting to preside this jury. I appreciated their interest, constructive comments, and kind encouragements.

My deep thanks go to the friends, the lab mates and the office mates, who accompanied me in this long journey. I thank Subhash, Marco, Loig and Ejder for having been such good friends. I thank Karim for his friendship, and for the infinite amount of coffee and jokes we shared. I thank to Axel for being a great office mate and for having read and commented my thesis at light speed, and Apostolos for having read every single paper of mine, correcting and discussing even the theorems and their proofs. I thank Lorenzo and Antonio for their help in Thales when my French level was null and I could not understand anyone else around me. I also thank Banhu, Gil, Mathieu, Salam, Leonardo, Alonso, Antonia, Daniel, Nikos, Sylvain, Abla, Azary, and all those ones who have shared with me at least a small part of this journey.

My heartfelt thanks go to my family. I thank my parents for their everlasting support. A very special thank-you goes to Lana, for her caring, help and support, and for being there for me in good times and bad. This thesis is dedicated to them.

Abstract

Modern communication systems are characterized by an increasing need for self-configuring networks. In fact, in many practical cases, the presence of centralizing devices such as a base station is neither realistic nor practical. This is the case, for instance, in military or emergency situation, or when the increasingly dense deployment of access points makes a man-made planning unfeasible. As a consequence, problems like designing behavioral rules for devices (or groups of devices) on how to select their own transmit parameters naturally arise. In particular, self-configuring algorithms must be able to respond to the necessity of detecting, avoiding or reducing interference, thus maintaining a sufficient quality of the communications when no centralization is available, and with minimum information exchange and cooperation. Moreover, these algorithms must be able to cope with the variations of the transmission conditions due to fading, shadowing, mobility and to the change in other devices behavioral patterns eventually creating extra interference.

The goal of this thesis is to study the joint problem of channel selection and power control in the context of multiple-channel clustered ad-hoc networks, i.e., decentralized networks in which radio devices are arranged into groups known as clusters, and to propose a viable decentralized self-configuring algorithm for such a network.

The network is studied and analyzed through game theory, and the relative equilibria are identified. The first purpose is to use these equilibria in order to quantify the performance of different algorithms that originate from the theory of learning in games. An algorithm based on the trial and error paradigm is then selected as a candidate solution. A particular utility function is designed in order for the equilibria to coincide with the solutions of an optimization problem, thus maximizing the quality of the communications while minimizing the resources needed. These results are presented in the most general form and therefore, they can also be seen as a framework for designing both games and learning algorithms with which decentralized networks can operate at global optimal points using only their available local knowledge.

The pertinence of the game design and the learning algorithm are highlighted using specific scenarios in decentralized clustered ad hoc networks. Numerical results confirm

the relevance of using appropriate utility functions and trial and error learning for enhancing the performance of decentralized networks.

Abstract (Français)

Les systèmes de communication modernes sont caractérisés par leur besoin croissant en mécanismes d'auto-configuration. En effet, dans de nombreux cas pratiques, la présence de dispositifs de centralisation tel qu'une station de base n'est ni réaliste ni pratique. Ceci est le cas, par exemple, des situations militaires ou aussi celles d'urgence, ou lorsque le déploiement de plus en plus dense de points d'accès rend la planification humaine irréalisable. Par conséquent, des problèmes tel que la conception de règles de comportement pour les appareils (ou groupes d'appareils) sur la façon de choisir leurs propres paramètres de transmission, se présentent naturellement. En particulier, les algorithmes d'auto-configuration doivent être en mesure de répondre à la nécessité de détecter, d'éviter ou de réduire les interférences, maintenant ainsi une qualité suffisante de communications quand une centralisation est indisponible, et ceci avec un minimum d'échange d'informations et de coopération. En outre, ces algorithmes doivent être en mesure de faire face aux variations naturelles des conditions d'émission, en raison de l'atténuation, des effets de masque, de la mobilité et de la variation des comportements des autres dispositifs qui peuvent éventuellement créer des interférences supplémentaires.

L'objectif de cette thèse est d'étudier le problème conjoint de sélection de canal et de contrôle de puissance dans le contexte de réseaux *ad hoc* clusterisés à canaux multiples, c'est à dire, des réseaux décentralisés dans lesquels les appareils radio sont disposés en groupes appelés clusters, et de proposer un algorithme d'auto-configuration décentralisé viable pour un tel réseau.

Le réseau est étudié et analysé par l'intermédiaire de la théorie des jeux, et les équilibres relatifs sont identifiés. Le premier objectif consiste à utiliser ces équilibres afin de quantifier les performances des différents algorithmes qui proviennent de la théorie de l'apprentissage dans les jeux. Un algorithme basé sur le paradigme "trial and error" est alors sélectionné en tant que solution candidat. Une fonction d'utilité particulière est conçue afin que l'équilibre puisse coïncider avec les solutions d' un problème d'optimisation, maximisant ainsi la qualité des communications, tout en minimisant les ressources nécessaires. Ces résultats sont présentés sous la forme la plus générale et, par conséquent, ils peuvent être aussi considérés comme un cadre théorique général pour la conception des jeux, ainsi que des algorithmes d'apprentissage avec lesquels les

réseaux décentralisés peuvent fonctionner à des points optimaux globaux, et ceci à l'aide uniquement de leurs connaissances locales disponibles.

La pertinence de la conception du jeu ainsi que de l'algorithme d'apprentissage est mis en évidence au moyen de scénarios spécifiques dans des réseaux *ad hoc* clusterisés et décentralisés. Les résultats numériques confirment la pertinence de l'utilisation des fonctions utilitaires appropriées ainsi que de l'apprentissage "trial and error" dans l'amélioration de la performance des réseaux décentralisés.

Acronyms

APC	average power consumption.
AS	average satisfaction.
BRD	best-response dynamics.
BS	base station.
BSMP	battle space spectrum management plan.
CCE	coarse correlated equilibrium.
CE	correlated equilibrium.
CH	cluster head.
CR	cognitive radio.
CRC	cyclic redundancy check.
CSMC	combined spectrum management cell.
CTFC	combined task force commander.
DO	distributed optimization.
DSCN	decentralized self-configuring *ad hoc* network.
DSG	dominant-solvable games.
DTMC	discrete time Markov chain.
ETE	enhanced trial and error.
FP	fictitious play.
GA	genetic algorithms.
GBDCA	greedy based decentralize control algorithm.
GT	game theory.
IG	interdependent game.
IP	internet protocol.
ISO	international standard organization.
IWF	iterative water-filling.
JUSTE-RL	joint utility and strategy estimation reinforcement learning.
LT	learning theory.
MAB	multi-armed bandit.
MAC	media access control.
MAI	multiple access interference.
MIMO	multiple-input multiple-output.

MNE	mixed Nash equilibrium.
NE	Nash equilibrium.
NET	network.
ODL	optimal dynamic learning.
PG	potential games.
PHY	physical.
QoS	quality of service.
RL	reinforcement learning.
RM	regret matching.
SC	small-cell.
SFP	smoothed fictitious play.
SINR	signal-to-interference-plus-noise ratio.
SMG	super modular games.
TE	trial and error.
UDP	user datagram protocol.
ZSG	zero sum games.

Résumé

Réseaux Décentralisés à Mécanismes d'Auto-Configuration

Au cours des dernières décennies, le trafic de données mobiles a tellement augmenté au point que l'infrastructure cellulaire standard n'est plus en mesure de faire face aux demandes de plus en plus croissantes. Smartphones, tablettes, ordinateurs et autres appareils portables ont renforcé le besoin de données sans fil et ont augmenté l'imprévisibilité de la demande de connectivité. L'architecture cellulaire standard est dépassée par la croissance de la demande de données réseau, d'où la nécessité de trouver de nouveaux paradigmes. Plusieurs améliorations technologiques ont été proposées afin d'éviter la congestion des réseaux cellulaires, tels que les multiple-input multiple-output (MIMO). L'une des solutions possibles est le déploiement dense de stations de base à courte distance appelées petites cellules (SCs) [1–3] en vue d'accroître considérablement la réutilisation spatiale. Ces SCs sont envisagées pour être facilement déployées et capables de détecter le spectre et de trouver de manière autonome la meilleure configuration de transmission. Un réseau composé de SCs et dispositifs similaires doit être capable de s'auto-configurer et prend le nom de réseau décentralisé auto-configurant (DSCN). Plus précisément, un DSCN est un réseau sans infrastructure dans lequel les émetteurs communiquent avec leurs récepteurs respectifs sans le contrôle d'une autorité centrale, par exemple, une station de base radio (BS). La pertinence de ces réseaux réside dans le fait qu'une planification formelle du réseau n'est pas nécessaire, que leur déploiement est simple, rapide et, plus important encore, des fonctionnalités telles que l'auto-guérison et l'auto-configuration sont souvent présentes. Par conséquent, les DSCNs couvrent un grand nombre d'applications, y compris militaires, de la police, de secours aux sinistrés, de l'espace ainsi que les applications commerciales indoor/outdoor [4, 5]. La technologie habilitante pour un tel réseau est la dite radio cognitive (CR) [6], un dispositif intelligent qui est capable d'observer son environnement et d'adapter ses paramètres de transmission afin d'optimiser ses fonctions objectives.

Les CRs et les DSCNs, d'abord envisagés pour augmenter le débit de données dans certaines applications civiles, jouent également un rôle important dans le domaine militaire, dans lequel l'obligation de garder le secret, la robustesse et l'adaptabilité des communications coïncident avec l'approche centralisée standard. Dans les champs hostiles

ou les situations d'urgence, une infrastructure de centralisation peut être peu pratique ou inefficace. Les utilisateurs malveillants pourraient exploiter le lien entre les stations de base et les appareils radio pour perturber ou intercepter la communication. En outre, un événement imprévisible est susceptible de perturber la communication entre la station radio base et le terminal, empêchant ainsi les communications. En revanche, les réseaux sans fils, au sein desquels les appareils sont en mesure de configurer leurs propres paramètres, sont intrinsèquement plus robustes à ce genre d'événements.

Un nombre croissant de recherches suggère que l'auto-organisation est l'une des capacités fondamentales que les réseaux décentralisés doivent présenter [7–12]. Le terme auto-configuration fait référence à la capacité des appareils radio de régler de manière autonome leur configuration d'émission-réception afin d'exploiter efficacement les ressources disponibles et garantir la fiabilité du réseau. Dans le cas le plus général, une configuration d'émission-réception peut être décrite en termes du nombre de bits d'information par bloc, de la longueur du bloc, du dictionnaire, des fonctions de codage-décodage, de la politique de sélection de canal, de la politique d'allocation de puissance, etc, comme suggéré dans [13–15].

Afin que les preneurs de décision dans les DSCNs puissent prendre des décisions efficaces, il faut qu'ils puissent s'appuyer sur des informations disponibles et fiables. La fonction de récupération de cette information prend le nom de détection [16]. De toute évidence, l'exécution de toute configuration sélectionnée augmente avec la précision et la fiabilité de l'information détectée. Théoriquement, s'il était possible de détecter tous les détails du réseau, un DSCN pourrait se configurer aussi bien qu'un réseau centralisé, avec un preneur de décision qui dispose d'informations complètes. Toutefois, les processus de détection présentent plusieurs problèmes [17, 18]. En effet, l'information acquise est toujours affectée par une certaine incertitude inhérente, l'effet de masque et l'affaiblissement étant inévitables. Pour surmonter cette limitation, un niveau croissant de coopération et d'échange d'informations entre les preneurs de décision doit être mis en place. Par conséquent, l'augmentation de la fiabilité de l'information obtenue ne peut être atteinte qu'aux dépends de la performance et de la sécurité. Par conséquent, les algorithmes d'auto-configuration qui visent à être implémentés de manière réaliste, devraient s'appuyer sur le minimum d'informations détectées possible. Dans ce contexte, la problématique de conception efficace d'algorithmes d'auto-configuration ainsi que la détermination des limites des DSCNs surgissent naturellement. Dans ce qui suit, une présentation de l'état de l'art des communications militaires ainsi que quelques-uns des principaux parcours de recherche concernant la conception des algorithmes dans les DSCNs.

Réseaux Militaires

L'état de l'art des communications sans fils dans les réseaux militaires repose sur un paradigme statique et centralisé. En règle générale, chaque réseau militaire (qui correspond normalement à une force militaire de la nation) est attribué à une fraction particulière du spectre. Lorsque les forces militaires doivent être déployées dans des zones hostiles, une phase de préparation de la mission a lieu, pendant laquelle chaque troupe se voit attribuer un canal logique particulier, par exemple une séquence de sauts notamment en mode de sauts de fréquence.

De manière plus détaillée, les techniques de gestion des ressources militaires non cognitives existantes sont définies dans [19], et identifient trois phases: la planification, le déploiement et la reconstruction.

- Phase de Planification

La phase de planification a pour but de créer le plan de gestion du spectre de l'espace bataille (BSMP). Le BSMP est un mappage entre les réseaux de dispositifs radio et la partie du spectre qui peut être exploitée. Cette cartographie comprend les tables d'affectation de canaux pour tous les équipements de la force de la coalition ainsi que les contraintes relatives à l'utilisation du spectre, tels que la puissance maximale d'émission, la hauteur de l'antenne et les zones de transmissions disponibles. Afin de créer le BSMP, une hiérarchie pyramidale des pouvoirs est mise en œuvre. Le commandant de la force de travail combinée dit (CTFC) désigne une cellule de gestion de spectre combiné (CSMC), qui remplit la tâche d'organiser et de coordonner les besoins en fréquences. À son tour, la CSMC établit un groupe de gestionnaires de composants de spectre qui, en général, représente les différents types de divisions militaires tels que la marine, l'infanterie, et l'aviation. Plusieurs pays de la coalition peuvent être présents dans chaque gestionnaire de composants de spectre. Par conséquent, chacune de ces nations est alors responsable de la création de son propre gestionnaire national du spectre qui organise la cartographie intra-national entre les appareils, l'utilisation du spectre et les zones opérationnelles relatives. En combinant les besoins en spectre de chaque nation, le CTFC compile l'ordre électronique de bataille qui détermine les missions du spectre complet.

- Phase de Déploiement

Durant cette phase, la force de chaque nation met en œuvre la disposition de la BSM, et observe le niveau global d' interférences. Les interférences peuvent avoir des origines amicales ou malveillantes. Elles sont donc soit le résultat d'une organisation incorrecte du spectre, soit d'une source de brouillage ennemie. Dans le cas où le niveau d'interférence

est élevé, la nation le reporte au niveau hiérarchique le plus haut, la CSMC. En collectant toutes les collisions et les niveaux d'interférence, la CSMC a pour mission de réduire les interférences en réorganisant opportunément les affectations de tables de fréquence, réduisant ainsi la puissance d'émission ou les attributions de créneaux horaires.

- Phase de Récupération

Durant cette phase, chaque nation informe les niveaux hiérarchiques plus élevés de l'instant auquel la fréquence assignée est restituée. En outre, si de nouvelles forces se joignent à la coalition, ou si de nouvelles exigences de spectre s'avèrent nécessaires pour les forces présentes, la CSMC compile un nouveau BSMP qui répond aux nouvelles exigences. D'après ce précède, il apparaît clairement que l'état de l'art des communications militaires présente plusieurs limitations, toutes liées à leur nature statique et fortement hiérarchique. En effet, les activités de gestion de fréquences qui sont effectuées lors de la phase de planification sont complexes et fastidieuses, en particulier dans les grandes coalitions. Par conséquent, il y a peu d'incitation à réorganiser les tâches une fois ces dernières sont déjà établies. Cela signifie que, de manière générale, une fois les mappings sont fixés, ils restent intacts pendant toute la durée de l'opération. D'autre part, une correspondance fixe entre les portions du spectre et les groupes de dispositifs, gaspille une grande partie du spectre qui pourrait rester inutilisée. En plus, cette correspondance fixe manque de flexibilité. Elle est donc peu pratique dans les cas où une partie des dispositifs est obligée de s'écarter des autres. En outre, ceci présente de sérieuses vulnérabilités aux sources de brouillage ennemies vu que les brouilleurs doivent concentrer leurs efforts uniquement sur une partie particulière du spectre [20]. Par conséquent, les communications de guerre modernes voient un intérêt croissant pour les DSCNs et les CRs [21–23], vu que la gestion dynamique du spectre peut éventuellement améliorer les performances et la sécurité des communications militaires, ce qui réduit également le nombre de niveaux hiérarchiques ainsi que la perte de flexibilité.

Allocation de Ressources Décentralisée

Compte tenu de l'intérêt de l'allocation de ressources de manière décentralisée ou distribuée, plusieurs cadres théoriques ont été développés pour tenter de trouver un système efficace [24–26]. Dans ce qui suit, une brève liste de solutions possibles est présentée et discutée.

Optimisation Distribuée

La théorie d'optimisation [27] est un outil mathématique qui vise à déterminer le maximum (ou le minimum) d'une fonction objective, sous certaines contraintes. Afin de mettre en œuvre la solution optimale (ou une solution sous-optimale) de manière décentralisée,

un cadre théorique nommé optimisation distribuée (DO) [28] a été développé. Basé sur la nature spécifique des fonctions objectives et des contraintes, le DO tente de diviser le problème en plusieurs sous-problèmes localement solvables [29–31]. Ces sous-problèmes, à leur tour, sont répartis entre une multitude de preneurs de décisions. Toutefois, l'obtention de sous-problèmes qui sont entièrement localement solvables est une tâche compliquée. Un certain niveau de collaboration, d'échange d'informations ou de synchronisation entre les différents preneurs de décision est obligatoire, voir par exemple [32].

Algorithmes Génétiques

Les algorithmes génétiques (GAs) [33] constituent une classe d'heuristiques basée sur le concept de l'informatique évolutionniste [34], et qui vise à déterminer le maximum de fonctions objectives multi-variables par l'intermédiaire de mécanismes imitant la sélection naturelle de gènes. Introduits dans le domaine de l'intelligence artificielle, les GAs représentent une classe d'algorithmes à convergence rapide qui se démarquent particulièrement dans les cas où la solution doit être choisie parmi un large ensemble. L'idée de base derrière les GAs consiste à créer un ensemble de codes génétiques, typiquement des séquences binaires représentant un des éléments possibles du domaine des fonctions objectives, pour ensuite les sélectionner à travers des mécanismes de sélection, de variation et d'héritage. Cependant, les GAs ne convergent pas nécessairement vers une solution optimale et leur mise en œuvre de manière complètement distribuée pose des problèmes non négligeables [35]. Afin d'allouer des ressources nécessaires pour les DSCNs, les solutions à base de GAs consistent à concevoir des fonctions ad hoc de fitness, qui sont maximisées par les preneurs de décision à travers un mécanisme de récompense [36]. Même si les GAs ont été mis en œuvre dans le but de configurer plusieurs paramètres en CR [37, 38], ces algorithmes nécessitent pour chaque radio d'avoir une vaste connaissance des règles de comportement des autres radios ainsi que leurs configurations possibles. Par exemple, toutes les CRs doivent partager une connaissance commune des mécanismes de récompense, des configurations disponibles et des paramètres réellement sélectionnés [39].

Théorie des Graphes

La théorie des graphes est un outil mathématique qui modélise les relations entre entités paires grâce à l'utilisation de structures mathématiques spécifiques appelées graphes [40, 41]. Les graphes sont faits de sommets, également connus sous le nom de nœuds, et des lignes qui les relient, appelés arcs. En appliquant cette théorie sur l'allocation des ressources dans les DSCNs, les nœuds représentent généralement les preneurs de décision (appareils, cellules, points d'accès). Deux nœuds sont reliés par un arc dans le cas où ils ne peuvent pas transmettre simultanément sur la même partie du spectre.

Dans ce cas, le problème d'allocation est réduit à un soi-disant problème de coloration de graphe [42]. Le problème de coloration de graphe est la tâche consistant à assigner des couleurs aux différents sommets d'un graphe de manière à ce que deux sommets adjacents soient affectés de couleurs différentes. Chaque couleur représente un canal particulier. Par conséquent, la résolution du problème de coloration de graphe revient à éviter les collisions possibles dans le réseau. Afin de trouver une solution au problème de coloration de graphe d'une manière dynamique et décentralisée, plusieurs techniques ont été proposées, voir par exemple [43–47]. Cependant, ces approches souffrent du défaut d'allouer uniquement les canaux, laissant ainsi sans solution le problème de configuration d'autres paramètres tels que la puissance.

Bandit Manchot

Le bandit Manchot (MAB) ("multi-arm bandit", en anglais) est un dilemme probabiliste auquel fait face un joueur quand il doit décider entre plusieurs machines à sous et qu'il a besoin de minimiser son regret [48, 49], c'est à dire, minimiser la perte due à une sélection non-optimale. Le joueur a besoin de concevoir une politique afin de choisir entre les actions qui apportent une forte récompense immédiate ou celles qui pourraient apporter une récompense plus élevée mais de façon non immédiate. Cette politique est composée d' une fonction d' indexage et d'une stratégie de sélection. La fonction d'indexation évalue la probabilité qu'une action particulière apporte une grande récompense, tandis que la stratégie de sélection décide, sur la base de l'indice, quelle machine doit être sélectionnée. En général, dans les DSCNs, les machines représentent les canaux logiques possibles, et les récompenses sont positives en cas de transmission réussie ou négatives en cas de collisions. Plusieurs fonctions d'indexation (par exemple, Gittins [50]) et politiques de sélection ont été proposées pour le MAB avec différents niveaux de raffinements et de détails [51–53] et avec des performances différentes [54–57]. Une des limitations de base de l'approche MAB repose sur le fait que le nombre de machines doit être supérieur au nombre de joueurs. Traduit dans une perspective radio, cela signifie que le nombre de canaux disponibles doit être supérieur au nombre de dispositifs potentiels [56, 58], ce qui rend l'approche MAB peu pratique dans les réseaux à densité élevée.

Théorie des Jeux et Théorie de L'apprentissage

La théorie des jeux (GT) est un cadre mathématique, né dans le domaine de l'économie [59], qui étudie les interactions stratégiques entre preneurs de décision rationnels concurrents appelés joueurs. D'une manière générale, la GT peut être divisée en GT coopérative, dans laquelle les joueurs sont libres de former des coalitions pour atteindre un objectif commun, et en GT non coopérative, dans laquelle les joueurs s'affrontent l'un contre l'autre pour atteindre un but égoïste [60, 61]. En GT non coopérative, le

concept de solution le plus largement utilisé est la célèbre notion de l'équilibre de Nash (NE) [62] et ses raffinements. Un NE est un état d'équilibre dans lequel aucun joueur ne peut améliorer son utilité par une déviation unilatérale. Lorsqu'elle est appliquée aux communications radio et aux problèmes d'allocation de ressources, la GT consiste à déterminer les limites de certaines solutions architecturales en étudiant les différentes solutions d'équilibres [63–73].

La motivation traditionnelle derrière la problématique de recherches d' équilibres revient au fait que ces derniers résultent naturellement de l' analyse des joueurs dans des situations où les règles du jeu, la rationalité des joueurs, ainsi que les fonctions de paiement des joueurs sont toutes connues [74]. Même si cette hypothèse semble cohérente avec l'observation empirique dans certains domaines, l'application d'un tel principe dans l'ingénierie radio semble irréalisable. Par conséquent, la détermination de procédures et d'algorithmes permettant aux réseaux de réaliser et de mettre en œuvre un équilibre, au moins sur une base stochastique, ou sur un équilibre approximatif, est encore un sujet ouvert [75, 76]. La théorie de l'apprentissage (LT) [9, 74, 77] est une tentative de conception d'algorithmes d'usage général permettant aux joueurs de mettre en œuvre différents types d'équilibres [7]. Ici, le terme usage général fait référence au fait que les algorithmes ne sont pas intrinsèquement liés à la nature des paramètres configurés, mais plutôt au jeu et à l'équilibre. Même si plusieurs algorithmes et programmes d'apprentissage différents ont été proposés afin de permettre aux réseaux de configurer leurs paramètres de transmission, par exemple, [7, 78–81], un cadre général pour mettre en œuvre tout NE de manière décentralisée et distribuée n'est pas encore défini.

Water-Filling Itératif

Le Water-Filling itératif (IWF) est probablement l'approche la plus largement étudiée pour l'allocation des ressources spectrales dans les DSCNs. Parmi les contributions les plus pertinentes concernant l'IWF, nous soulignons celles de [79, 82–88]. Même si l'IWF peut être considéré comme un cas particulier de l'algorithme d'apprentissage "best-response dynamics" (BRD), cette approche est considérée comme étant originaire du domaine de la théorie de l'information [88], et ses applications ont été d'abord étudiées dans le cadre du contrôle de puissance de lignes d'abonnés numériques. En bref, l'IWF permet à chaque émetteur autonome de diviser sa puissance disponible sur tous les canaux de transmission, Water-Filling par rapport aux gains du canal et aux niveaux du rapport signal-sur-interférence-plus-bruit (SINR). Puisque la solution du Water-Filling est connue pour fournir l'efficacité spectrale la plus élevée sur les liens à entrée et sortie uniques (SISO) [89], elle apparaît comme une solution naturelle pour maximiser également l'efficacité spectrale dans les réseaux à accès multiples. Cet algorithme peut être utilisé avec succès à la fois pour optimiser le débit sous une contrainte de puissance d'émission maximale [90] et pour réduire au minimum la puissance utilisée, tout en réalisant un taux de transmission cible [85].

Cependant, cette approche présente deux problèmes principaux. Tout d'abord, la preuve de convergence de cet algorithme [79, 91] est soumise à l'hypothèse sous laquelle le système fonctionne dans le régime d'interférences faibles. Ensuite, il existe une littérature suffisante [67, 92, 93] qui démontre que dans les DSCNs, le point d'opération obtenue par IWF est souvent inefficace.

Autres Techniques

Il existe de nombreuses autres techniques permettant de concevoir des algorithmes d'auto-configuration. En plus, il existe une vaste littérature (par exemple, [94–99]), de techniques heuristiques permettant de configurer certains paramètres particuliers dans un scénario particulier. Par exemple, dans [43], un réseau à canaux multiples ad hoc clusterisé dans lequel les clusters sont en mesure de détecter tous les canaux disponibles est considéré. Quand un canal sans interférence n'est pas disponible, le choix du canal est fait au hasard. Dans les réseaux à faible densité de population, cette règle de comportement présente une performance acceptable avec très peu de complexité de mise en œuvre. Néanmoins, dans les réseaux à haute densité démographique, cette approche s'avère très sous-optimale. Cependant, l'origine de ces techniques basée sur l'expérience ne permet pas une analyse théorique. Leur performance ne peut donc être évaluée et comparée qu'à partir uniquement d'exemple réels ou de séries de simulation. En outre, le manque de compréhension théorique les rend non adaptées à la configuration de différents types de paramètres.

Observations Finales

D'après ce qui précède, il est clair qu'il existe de nombreuses options viables permettant de concevoir des algorithmes d'auto- configuration pour les DSCNs. Chaque approche est différente de l'autre par, en gros, trois caractéristiques: les hypothèses de l'information, donc ce dont chacun des algorithmes a besoin de savoir sur l'environnement, ainsi que le volume d'informations nécessaire à échanger entre les appareils; le type de solution mise en place et ses performances respectives; la possibilité d'analyser théoriquement le résultat de l'algorithme. Afin d'établir laquelle des approches précédentes correspond le mieux à un problème d'allocation de ressources donné, un cadre théorique commun de base doit être défini. Cette base pourrait être exploitée afin de comparer les performances et d'équilibrer les hypothèses de l'information ainsi que la pénibilité de calcul de chaque algorithme. Malheureusement, une véritable formation théorique est manquante, laissant simulations et prototypages comme seuls tests possibles. L'approche suivie dans cette thèse commence par la sélection de la théorie des jeux GT comme outil de description mathématique pour les DSCNs, et adopte ensuite un algorithme d'apprentissage permettant la conception d'un algorithme d'auto-configuration. Les principales motivations derrière sont les suivantes: (i) par rapport aux autres approches la GT et la

LT ne supposent aucun paramètre de configuration particulier, les algorithmes peuvent être donc conçus pour mettre en œuvre différentes configurations, (*ii*) l'analyse de l'équilibre est un outil d'étude perspicace des limites de la performance des DSCNs, (*iii*) les autres approches supposent souvent un certain niveau de coordination et d'échange d'informations entre les différents preneurs de décision comme une condition nécessaire.

Plan et Contributions

Cette thèse se compose de cinq chapitres différents: Introduction (chapitre 1), Théorie (chapitre 2), Modèle du système (chapitre 3), Applications et résultats (chapitre 4), Conclusions et Perspectives (chapitre 5).

Le chapitre 2 présente le contexte théorique de la thèse. Dans la Section 2.1, le jeu de notations théoriques, les concepts, ainsi que les principaux concepts d'équilibre existants pour les DSCNs sont introduits [7]. La Section 2.2 divise les algorithmes d'apprentissage en deux catégories: ceux à convergence asymptotique, et ceux basés sur le mécanisme du "trial and error". La convergence asymptotique de différents algorithmes d'apprentissage est comparée en termes de besoin en informations et de propriétés de convergence. Les limitations majeures de tels algorithmes réside d'abord, dans la nécessité d'avoir une structure particulière du jeu afin que la prédiction du résultat soit possible et que la convergence soit assurée, et ensuite du niveau d'informations important dont chaque joueur doit disposer afin que l'algorithme soit efficace. La SectionSection 2.2.4 présente l'algorithme d'apprentissage "trial and error" qui a prouvé son efficacité dans une vaste variété de jeux. Parmi les caractéristiques de cet algorithme, on note en particulier le Théorème 2.10 [100, 101], qui démontre sa convergence stochastique vers le NE qui maximise les performances du réseau.

Le chapitre 3 introduit et analyse un modèle abstrait d'un DSCN. Ce modèle peut être utilisé pour représenter à la fois, les réseaux sans fils militaires, et les réseaux civils. Nous supposons que le but du concepteur est de maximiser une fonction globale qui représente la qualité des communications dans le réseau comme le SINR ou le débit, tout en minimisant l'utilisation des ressources, par exemple l'épuisement de la batterie. En effet, de nombreuses applications réelles nécessitent un minimum de qualité de communications afin de fonctionner correctement. Par exemple, l'application vocale instantanée et les applications vidéo peuvent nécessiter un débit minimum, et leur qualité ne s'améliore pas radicalement une fois ce minimum est dépassé. D'autre part, la consommation de la batterie est un élément clé dans les communications mobiles. Il est donc nécessaire de réduire la consommation de puissance afin de réaliser des communications à long terme. De toute évidence, ceci est d'une importance vitale dans le domaine militaire ainsi que dans les situations d'urgence. Le chapitre 3 introduit également deux instances de DSCNs utilisées comme scénarios de tests pour les différents algorithmes: un

DSCN dense statique, et un DSCN mobile. Le modèle du DSCN basé sur la théorie des jeux est présenté dans la Section 3.3. Une fonction d'utilité particulière est définie afin d'atteindre le NE avec la fonction de bien-être social la plus élevée, et qui coïncide avec la solution du problème d'optimisation défini dans la Section 3.1. L'objectif consiste à exploiter la propriété de l'algorithme d'apprentissage TE de converger vers le NE avec la fonction de bien-être social la plus élevée afin de converger vers la solution du problème d'optimisation, et par conséquent configurer le réseau de manière optimale. Une des propriétés remarquable de cette fonction telle que, si chaque lien dans chaque cluster peut évaluer son propre QoS et transmettre un message de 1 bit au cluster head (CH), alors ce dernier est capable de calculer la valeur de cette utilité moyennant uniquement les informations intra-cluster.

Le chapitre 4 présente les principaux résultats de cette thèse. Dans la Section 4.1, le résultat théorique concernant l'algorithme TE est présenté. Les théorèmes 4.1 et 4.2 établissent un lien précis entre le NE avec le bien-être social le plus élevé, et le résultat du problème d'optimisation. Ce lien permet au concepteur du réseau de sélectionner l'objectif du réseau à travers la définition de la fonction objective et les différentes contraintes du problème d'optimisation défini dans la Section 3.1. Cette fonction d'utilité sera utilisé par l'algorithme TE afin de diriger le réseau vers la solution du problème d'optimisation.

Le Théorème 4.3 évalue les bornes supérieures et inférieures du nombre d'itérations moyen dont l'algorithme d'apprentissage TE a besoin avant d'atteindre le NE pour la première fois. Le Théorème 4.4 quant à lui, fournit une approximation de la fraction de temps au cours de laquelle l'algorithme joue un NE. Ces deux résultats, validés numériquement dans la Section 4.1.3, permettent de démontrer que les deux quantités dépendent du paramètre ε.

Les algorithmes présentés dans la Section 2.2 sont comparés en termes de performances dans la Section 4.2. Il est possible d'observer, qu'en général, les algorithmes nécessitant une plus grande quantité d'informations concernant la structure du jeu, atteignent des points d'opérations plus performants. D'autre part, il est démontré que ce genre d'algorithmes, subit une chute radicale de performances dans certains DSCNs. La Section 4.3 teste la performance de l'algorithme TE dans les scénarios statiques et mobiles définis dans la Section 3.1. Les limites de cet algorithme, notamment en termes d'instabilité et de politique d'expérimentation sous-optimale, sont identifiées. Une version améliorée de cet algorithme, visant à faire face à ces limitations, est implémentée et décrite dans la Section 4.5. Cette amélioration s'inspire de la théorie développée dans la Section4.1, afin d'identifier une solution au manque de stabilité de l'algorithme solution. Cette version améliorée est ensuite testée et validée, d'abord par comparaison avec un TE standard dans la Section 4.5.3, et ensuite avec d'autres algorithmes d'apprentissage dans la Section 4.5.4, ce qui permet de démontrer son efficacité dans la configuration des DSCNs.

Des résultats préliminaires, réalisés dans le cadre du projet CORASMA, et fournissant une première validation réelle de cette version, sont présentés dans la Section 4.5.5. Cette thèse est finalement conclue dans le chapitre 5, qui résume les principaux résultats et fournit une perspective sur de possibles travaux futurs. En particulier, les difficultés rencontrées dans des systèmes réels sont analysés.

Publications

Le travail présenté tout au long de cette thèse est le résultat de plusieurs publications et travaux menés par le programme CORASMA (radio cognitive pour la gestion dynamique du spectre) de l'Agence européenne de défense (EDA). Ce travail a déclenché divers brevets qui sont en cours de révision par le ministère français de la Défense. Les principaux résultats de cette thèse sont résumés dans les articles ci-dessous.

Articles de Revue

- L. Rose, L., S. Lasaulce, S. M. Perlaza, M. Debbah, Learning equilibria with partial information in decentralized wireless networks, *IEEE Communications Magazine*, Vol 49, no. 8, pp. 136–142, Aug. 2011.
- L. Rose, L., S. M. Perlaza, C. J. Le Martret, M. Debbah, Self-Organization in Decentralized Networks: A Trial and Error Learning Approach *accepted for publication on IEEE Transactions on Wireless Communications*, 2013.

Conférences Internationales avec Actés

- L. Rose, L., S. M. Perlaza, M. Debbah, On the Nash equilibria in decentralized parallel interference channels, *Proc. of IEEE Workshop on Game Theory and Resource Allocation for 4G*, Kyoto, Japon, pp. 1–6, Jun. 2011.
- L. Rose, S. M. Perlaza, M. Debbah, C. J. Le Martret, Distributed power allocation with SINR constraints using trial and error learning, *in Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, pp. 1835–1840, Apr. 2012.
- L. Rose, C. J. Le Martret, M. Debbah, Channel and power allocation algorithms for *ad hoc* clustered networks, *in Proc. of the Military Communications and Information Systems Conference (MCC)*, Gdansk, Poland, pp. 1–8, 8–9 Oct. 2012.
- L. Rose, E. V. Belmega, W. Saad, M. Debbah, Dynamic service selection games in heterogeneous small cell networks with multiple providers, *in Proc. of the International Symposium on Wireless Communication Systems (ISWCS)*, Paris, France, pp. 1078–1082, 28–31 Aug. 2012.

- L. Rose, S. M. Perlaza, C. J. Le Martret, M. Debbah, Achieving Pareto optimal equilibria in energy efficient clustered *ad hoc* networks, *the IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, pp. 1491–1495, 1–8 Jun. 2013.
- L. Rose, C. J. Le Martret, M. Debbah, Joint Channel and Power Allocation in Tactical Cognitive Networks: Enhanced Trial and Error, *in Proc. of the Military Communications and Information Systems Conference (MCC)*, Saint-Malo, France, pp. 1–11, 7–9 Oct. 2013.
- L.Rose, R. Massin, L. Vijayandran, M. Debbah, C. J. Le Martret, CORASMA Program on Cognitive Radio for Tactical Networks: High Fidelity Simulator and First Results on Dynamic Frequency Allocation, *to appear in Proc. of the IEEE Military Communications Conference (Milcom)*, San Diego, CA, USA, 18–20 Nov. 2013

Conclusions & Perspectives

Conclusions

Afin de faire face à la demande croissante en services de données sans fils, les nouveaux systèmes de communications mobiles seront constitués de stations de base à courte distance densément déployées, tels que les SCs ou encore les indoor Femtocells. Ces dispositifs sont envisagés afin d'être mis en œuvre avec un minimum de planification, et sont pensés pour être capables d'explorer en permanence leur environnement et d'adapter de manière optimale leurs caractéristiques. Par conséquent, les réseaux caractérisés par la présence massive de ces dispositifs auront un besoin croissant d'exploiter intelligemment les ressources disponibles, d'où un besoin grandissant pour des algorithmes efficaces en mesure de configurer de manière optimale les paramètres du réseau.

Dans les réseaux militaires et d'urgence, il est naturellement nécessaire de garantir le secret et la flexibilité des communications. Aujourd'hui, les communications militaires se présentent avec une gestion du spectre hiérarchisée de manière pyramidale ayant l'humain au centre des décisions et une gestion de ressources complètement centralisée. Dans ce contexte, la présence d'infrastructures de télécommunications fixes n'est ni pratique, ni souhaitable. Les contraintes évidentes en raison de la rudesse des conditions s'ajoutent aux points faibles que la présence d'une BS offre à un utilisateur malveillant, ce qui rend l'auto-configuration une fonctionnalité encore plus désirable. Atténuer les interférences, éviter les collisions, et réduire la consommation en énergie dans ces réseaux est donc de grande importance. Toutefois, en raison de l'imprévisibilité des conditions sans fils, des fonctionnalités d'auto-configuration deviennent une caractéristique

nécessaire. Un tel réseau complexe, composé de dispositifs d'auto-configuration intelligents, exige un nouveau cadre théorique pour analyser leurs performances.

Dans cette thèse, un modèle théorique du jeu dans les DSCNs est proposé et plusieurs algorithmes d'apprentissage pour les réseaux munis de mécanismes d'auto-configuration sont étudiés et discutés. La pertinence de ces algorithmes appliqués aux communications sans fil est identifiée en termes de contraintes pour le système (l'ensemble des actions continues ou discrètes, les informations requises, les hypothèses de l'information, la synchronisation, la signalisation, etc.), ainsi que les critères de performance tels que l'utilité obtenue à l'état d'équilibre, la vitesse de convergence, etc. Un lien précis entre les algorithmes et les concepts d'équilibre concernés est établi. Ce lien pourrait permettre à un concepteur de réseau de définir l'ensemble des actions particulières et des fonctions d'utilité afin de permettre à l'équilibre d'avoir des caractéristiques particulièrement intéressantes, comme la haute équité, la performance globale élevée, etc. Les limites et les inconvénients de ces algorithmes sont évalués, et un algorithme en particulier, à savoir le "trial and error", est sélectionné afin de configurer un DSCN militaire. Les principales raisons en sont que les algorithmes d'apprentissage asymptotiques exigent une structure particulière du jeu afin de converger vers l'équilibre. D'une part, un modèle de jeu de type DSCN respecte rarement l'un de ces types. L'algorithme d'apprentissage TE s'avère donc un candidat convenable, vu sa capacité à converger dans différents jeux. Les particularités de cet algorithme peuvent se résumer comme suit:

- Il est constitué d'une machine d'état qui s'exécute à chaque preneur de décision;
- Il requiert une connaissance minimale sur le jeu joué;
- Il ne nécessite qu'une estimation numérique de l'utilitaire à chaque itération;
- Ses états sont stochastiquement stables, les états qui sont les plus susceptibles d'être joués dans le long terme, sont les équilibres de Nash réalisant le bien-être social le plus élevé;
- Il a besoin d'une réinitialisation et il répond rapidement aux changements survenant dans le réseau, grâce à l'absence d'un état de convergence asymptotique.

Les principaux problèmes associés à cet algorithme ont été identifiés dans l'instabilité de l'association canal-cluster ainsi que dans la politique de sélection de configuration sous-optimale. Pour offrir une solution à ces problèmes, une nouvelle version améliorée de l'algorithme est développée. Ses principales caractéristiques sont basées sur la présence du facteur d'expérimentation double, et sur une politique permettant un choix de configuration plus intelligent qui teste uniquement les configurations qui peuvent être optimaux. Le rôle de chaque nouveau paramètre est discuté et son effet sur les capacités de convergence de l'algorithme est évalué. En particulier, on remarque qu'une

faible fréquence d'expérimentation dans l'association canal -cluster est recommandée dans le réseau statique, tandis que la présence d'événements imprévisibles, tel que l'affaiblissement, accroissent la nécessité d'une réponse rapide. Les capacités de cet algorithme sont ensuite évaluées selon divers scénarios statiques et/ou mobiles en présence et/ou absence de canaux à évanouissement.

Malgré les performances remarquables dans la configuration des DSCNs dont a fait preuve cet algorithme, quelques mots de mise en garde sont nécessaires en ce qui concerne nos résultats. Dans les réseaux réels, des événements imprévisibles peuvent aussi provenir de l'intérieur du cluster. Les appareils qui souhaitent ne pas transmettre, les radios qui ne sont plus fonctionnels, une évaluation erronée de la QoS perçue, ou encore un retour corrompu, peuvent détériorer la fiabilité de l'estimation de la fonction d'utilité au sein du CH.

Perspectives

Différentes perspectives peuvent être envisagées comme extensions possibles au travail effectué au cours de cette thèse.

Modélisation Décentralisée des Réseaux Auto-Configurés

La GT s'est révélée être un outil puissant de modélisation des DSCNs. La liste croissante de raffinages mathématiques de la théorie comme les jeux stochastiques, visent à améliorer la qualité de modélisation du comportement d'un joueur indépendant dans un environnement réel. Cependant, même ces typologies de jeux échouent dans la modélisation exacte d'un DSCN. Par exemple, le positionnement des dispositifs ainsi que leur apparition ou disparition sont rarement modélisés. Une contribution pertinente à la résolution de ce problème peut provenir de la géométrie stochastique qui offre un cadre mathématique de développement de modèles de réseaux pour lesquels les emplacements des dispositifs, et la structure du réseau sont des variables aléatoires. Une caractérisation mathématique complète des DSCNs pourrait améliorer la compréhension de son mécanisme et conduire à de meilleurs algorithmes d'auto-configuration.

Conception d'Algorithmes

Comme déjà mentionné, plusieurs approches algorithmiques alternatives aux algorithmes d'apprentissage existent dans la littérature. Une procédure itérative capable d'apprendre un équilibre particulier ou qui montre un résultat prévisible et évaluable dans un large éventail de cas est cependant absente. Les inégalités variationnelles sont de plus en plus perçues comme permettant d'atteindre des états stables et prévisibles dans un large

ensemble de typologies de scénarios. En outre, des algorithmes permettant de mieux traiter les informations de détection et peut-être même de déclencher la détection de paramètres particuliers ouvrent un chemin viable et intéressant à l'amélioration de la qualité des configurations.

Théorie des Jeux

Le rôle de la GT est loin d'être complètement déterminé dans le domaine de DSCNs. Comme déjà démontré, un certain niveau de centralisation persiste dans les réseaux réels, même s'il ne s'agit que d'une centralisation locale. Un développement possible dans ce sens pourrait consister en l'étude de la centralisation locale à travers une GT coopérative. Cela pourrait conduire à des algorithmes de sélection de CH plus efficaces et plus dynamiques ou à un algorithme de contrôle décentralisé qui permet au réseau d'obtenir le même résultat que celui centralisé et ceci sans avoir besoin d'un contrôleur central.

Algorithme "Trial and Error"

Comme mentionné précédemment, l'algorithme développé dans cette thèse ne tient pas compte des modifications intra-cluster. En outre, son comportement en cas de division ou de fusion du groupement doit encore être analysé. Les travaux dans ce sens permettraient une compréhension englobant tous les événements possibles tels que l'apparition ou la disparition de nœuds ou encore une interprétation poussée des erreurs dans les évaluations. En outre, une politique d'expérimentation optimale pourrait être développée afin d'améliorer la stabilité ainsi que les performances. La mise en œuvre de tests s'appuyant sur des simulateurs Hi-Fi et leur prototypage est également un développement intéressant qui pourrait permettre d'évaluer les limites de l'algorithme dans des conditions réelles.

Chapter 1

Introduction

1.1 Decentralized Self Configuring Networks

In the late decades, mobile data traffic has risen up to the point that standard cellular infrastructure is not able to cope with growing demands. Smartphones, tablets, laptop PCs and other portable devices have boosted the wireless data need and increased the unpredictability of connectivity demand. Standard cellular architecture development is outpaced by the growing network data demand, thus new paradigms have to be found. Several technological improvements have been proposed in order to avoid the congestion of cellular networks, such as multiple-input multiple-output (MIMO). One of the possible solutions is the dense deployment of short ranged base stations known as small-cells (SCs) [1–3] in order to drastically increase the spatial reuse. These SCs are envisioned to be easily deployable and able to sense the spectrum and autonomously find the best transmission configuration. A network composed by SCs and similar devices needs to be able to self configure and takes the name of decentralized self-configuring *ad hoc* network (DSCN). More precisely, a DSCN is an infrastructure-less network in which transmitters communicate with their respective receivers without the control of a central authority, for instance, a base station (BS) . The relevance of these networks lies in the fact that a formal network planning is not required, their deployment is easy, quick and, more importantly, capabilities such as self-healing and self-configuration are often present. Therefore, DSCNs span a large number of applications including military, law enforcement, disaster relief, space, and indoor/outdoor commercial applications [4, 5]. The enabling technology for such a network is the so called cognitive radio (CR) [6], an intelligent device that is able to observe its environment and adapts its transmission parameters in order to optimize its objective functions.

CRs and DSCNs, first envisioned for increasing the data rate in civilian applications, also play a relevant role in the military field, in which the requirement for secrecy, robustness and adaptability of communications collides with the standard centralized

approach. In hostile or emergency fields, a centralizing infrastructure can be either unpractical or inefficient. Malicious users could exploit the link between BSs and the radio devices to disturb or eavesdrop the communication. Moreover, an unpredictable event may disrupt communication between the BS and the end terminal, thus preventing the communications. On the contrary, wireless networks in which all devices are able to self configure their parameters are inherently more robust to these events.

A growing body of research suggests that self-organization is one of the fundamental capabilities decentralized networks must exhibit [7–12]. The term self-configuration refers to the capability of radio devices to autonomously tune their transmit-receive configuration for efficiently exploiting the available resources and guaranteeing network reliability. In the most general case, a transmit-receive configuration can be described in terms of the number of information bits per block, the block length, the codebook, the encoding-decoding functions, the channel selection policy, the power allocation policy, etc., as suggested in [13–15].

In order for the decision takers in DSCNs to take efficient decisions, it is necessary that they can rely on available and reliable information. The function of retrieving this information takes the name of sensing [16]. Clearly, the performance of any selected configuration increases with the precision and reliability of the sensed information. Theoretically, if it were possible to sense all the details of the network, a DSCN could configure itself just as well as a centralized network, with a decision taker with full information. However, sensing processes present several problems [17, 18]. The information acquired is always affected by an inherent uncertainty, as shadowing and fading are unavoidable. To overcome this limit, a growing level of cooperation and information exchange among the decision takers must be put in place. Therefore, increasing the reliability of the sensed information can be achieved only at the expenses of performance and security. As a result, self-configuring algorithms that aim at being realistically implemented should rely on the minimum possible sensed information.

In this context, the problematics of how to efficiently design self-configuring algorithms and what are the limits of DSCNs have arisen naturally. The following presents the state of the art in military communications and introduces some of the main research path followed for algorithm designing in DSCNs .

1.1.1 Military networks

The state of the art wireless communications in military networks is based on a static and centralized paradigm. In general, each military network (which normally corresponds to one nation's military forces) is assigned with a particular fraction of the spectrum. When the military forces need to be deployed into unfriendly zones, a mission preparation phase

takes place, in which each troop is assigned a particular logical channel, for instance a particular sequence of hops in a frequency hop fashion.

More in detail, existing non-cognitive military resource management are defined in [19], which identifies three phases: planning, deployment and recovery.

- Planning phase

The planning phase has the purpose of creating the battle space spectrum management plan (BSMP). The BSMP is a mapping between radio devices' networks and the portion of the spectrum that can be exploited. This mapping includes the channel assignment tables for all equipments in the coalition force and the constraints on the use of the spectrum, such as maximum transmit power, antenna height and available zones of transmissions. In order to create the BSMP, a pyramidal hierarchy of authorities is implemented. The so called combined task force commander (CTFC) nominates a combined spectrum management cell (CSMC), which fulfills the task of organizing and coordinating the spectrum requirements. In turn, the CSMC establishes a group of component spectrum manager that, generally, represents different types of military divisions such as navy, infantry, aviation. Several of the coalition's nations can be present in each each component spectrum manager. Therefore, each of these nations is then responsible of creating its own national spectrum manager that organizes the intra-national mapping between devices, spectrum usage and the relative operational areas. By combining each nation's spectrum need, the CTFC compiles the electronic order of battle that determines the complete spectrum assignments.

- Deployment phase

During this phase, each nation's force implements the disposition of the BSMP, and observes the overall level of interference. Interference may have both friendly or malicious origins, that is, it may originate by a non-correct organization (or implementation) of the spectrum, or by an enemy's jamming source. In case of an elevated interference level, the nation reports to the higher hierarchical level, the CSMC. Collecting all the collisions and interference levels, the CSMC has the task of alleviating interference disturbance by opportunely reorganizing the frequency table assignments, thus reducing the transmission power or the time slot assignments.

- Recovery phase

During this phase, each nation informs the higher hierarchical levels of the instant in which the assigned frequency is handed back. Moreover, if new forces join the coalition,

or new spectrum demands are made necessary for the present forces, the CSMC compiles a new BSMP meeting the new requirements.

From the description above, it becomes clear that state of the art military communications presents several drawbacks, all linked with their static and overly hierarchical nature. In fact, frequency management activities that are performed during the planning phase are complex and time consuming, especially in large coalitions. Hence, there is little incentive to reorganize the assignments once they are made. This means that, generally, once the mappings are fixed, they remain untouched for the whole duration of the operation. On the other hand, a fixed correspondence between spectrum's portions and groups of devices wastes large parts of the spectrum that might remain unused. Furthermore, this fixed correspondence lacks of flexibility, thus it is unpractical in cases in which a part of the devices is forced to depart from the others. Furthermore, it presents serious vulnerabilities to enemies' jammers or eavesdroppers that need to focus their efforts only on a particular portion of the spectrum [20]. Therefore, modern warfare communications are seeing an increased interest in DSCNs and CRs [21–23] as dynamic spectrum management can possibly improve both the performance and the security of military communication, also reducing the amount of hierarchical level and consequent loss of flexibility.

1.1.2 Decentralized resource allocation overview

Given the interest in allocating the resources in a decentralized or distributed way, several theoretical frameworks have been developed to attempt to find an efficient scheme [24–26]. In the following, a short list of possible approaches is presented and discussed.

1.1.2.1 Distributed Optimization

Optimization theory [27] is a mathematical tool that aims at finding the maximum (or the minimum) of an objective function under certain constraints. In order to implement the optimal (or a suboptimal) solution in a decentralized way, a theoretical framework named distributed optimization (DO) [28] has been developed. Based on the specific nature of the objective functions and constraints, DO attempts to divide the problems into locally solvable subproblems [29–31]. These subproblems, in turn, are distributed among a multitude of decision-takers. However, obtaining subproblems that are fully locally solvable is a complicated task, thus often a certain level of collaboration, information exchange or synchronization among the different decision-takers is mandatory, see for instance [32].

1.1.2.2 Genetic Algorithms

Genetic algorithms (GA) [33] are a class of heuristics based on the concept of evolutionary computing [34] that aim at finding the maximum of multi-variable objective functions through mechanisms that mimics the natural selection of genes. Introduced in the field of artificial intelligence, GAs are a class of fast converging algorithms that performs particularly well in cases in which the solution must be chosen from a large set. The basic idea behind GAs is to create a set of genetic codes, typically binary strings representing one of the possible elements of the domain of objective functions, and then selecting them through the mechanisms of selection, variation and inheritance. However, GAs do not need to converge to an optimal solution and their implementation in a completely distributed way poses non-trivial problems [35]. In order to allocate resources for DSCNs , GAs based solutions consist in designing *ad hoc* fitness functions, that are maximized by the decision takers through a reward mechanism [36]. Even though GAs have been implemented to configure several parameters in CRs [37, 38], these algorithms require for each radio to have a vast knowledge on the other radios behavioral rules and possible configurations. For instance, all CRs need to share a common knowledge of the reward mechanisms, of the available configuration and of the parameters actually selected [39].

1.1.2.3 Graph Theory

Graph theory is a mathematical tool that models pairwise relations between entities through the use of particular mathematical structures known as graphs [40, 41]. Graphs are made of vertices, also known as nodes, and lines connecting them, known as edges. When applied to resource allocation in DSCNs, nodes usually represent the decision takers (devices, cells, access points). Two nodes are connected by an edge in the case in which they cannot simultaneously transmit on the same spectrum portion. In this case, the allocation problem reduces to a so called graph coloring problem [42]. The graph coloring problem is the task of assigning colors to the vertices of a graph in such a way that two adjacent vertices are assigned different colors. Each color represents a particular channel, therefore solving the graph coloring problem coincides with avoiding possible collisions in the network. In order to find a solution of the graph coloring problem in a dynamic and decentralized way, several techniques have been proposed, see among the others [43–47]. However, such approaches suffer from the defect of allocating only the channels, leaving unsolved the problem of configuring other parameters such as transmit power.

1.1.2.4 Multi-armed Bandit Theory

The multi-armed bandit (MAB) is a probabilistic dilemma that gamblers face when they have to decide between several slot machines (each known as the one-armed bandit) and they need to minimize their regret [48, 49], i.e., minimizing the loss due to selecting non-optimally. The gambler needs to design a policy in order to choose between actions that bring an immediate high reward or actions that might bring a higher but late reward. This policy is composed of an indexing function and a selection strategy. The indexing function evaluates the probability of a particular action of bringing a high reward, while the selection strategy decides, based on the index, which arm has to be selected.

Generally, when applied to DSCNs, the arms represent the possible logical channels, and the rewards are positive in case of successful transmission or negative in case of collision. Several indexing functions (e.g., Gittins [50]) and selection policies have been proposed for the MAB with different levels of refinement and detail [51–53] and with different performance [54–57]. One of the basic limits of the MAB approach relies on the fact that the number of arms must be greater than the number of gamblers. Translated into a radio perspective, this means that the number of available channels must be greater than the number of potential devices [56, 58], thus making the MAB approach unpractical in dense networks.

1.1.2.5 Game Theory and Learning Theory

Game theory (GT) is a mathematical framework, born in the field of economics [59], that investigates the strategical interactions between competing, rational decision takers known as players. Broadly speaking, GT can be divided into cooperative GT, in which players are free to form coalitions to achieve a common goal, and non-cooperative GT, in which each player competes with each other to achieve a selfish goal [60, 61]. In non-cooperative GT, the most widely used solution concept is the celebrated notion of Nash equilibrium (NE) [62] and its refinements. A NE is an equilibrium state of the game in which no player can improve its utility by a unilateral deviation.

When applied to radio communication and resource allocation problems, the role of GT is to determine the limits of certain architectural solutions by studying the various equilibria solutions [63–73].

The traditional motivation for when and why equilibria arise is that they naturally result from the analysis of the players in situations where the rules of the game, the rationality of the players, and the players' payoff functions are all common knowledge [74]. Even though this assumption seems consistent with the empirical observation in some fields, the applicability of such a principle in radio engineering seems unfeasible. As a result, determining procedures and algorithms in order to let networks achieve

and implement an equilibrium, at least on stochastic basis, or an approximation of an equilibrium, is still an open problem [75, 76]. Learning theory (LT) [9, 74, 77] is an attempt to design general purpose algorithms to allow players to implement different kind of equilibria [7]. Here, the term general purpose refers to the fact that the algorithms are not inherently linked with the nature of the parameters that are configured, rather with particular the game and with the equilibrium. Even though several different algorithms and learning schemes have been proposed in order to enable networks self configure their transmission parameters, e.g. [7, 78–81], a general framework to implement any NE in a decentralized and distributed way is still missing.

1.1.2.6 Iterative Water-Filling

Iterative water-filling (IWF) is probably the most widely studied approach for allocating spectral resources in DSCNs. Among the most relevant contributions regarding the IWF , we highlight those in [79, 82–88]. Even though IWF can be considered a special case of a learning algorithm known as best-response dynamics (BRD), it considered as originated in the field of information theory [88] and its applications were first studied in digital subscriber lines' power control. Briefly, the IWF lets each transmitter autonomously divide its available power among all the transmission channels, water-filling with respect to the channel gains and the signal-to-interference-plus-noise ratio (SINR) levels. Since the water-filling solution is known to provide the highest spectral efficiency in single-input single-output links [89], it appears as a natural solution to maximize also multiple access networks spectral efficiency. This algorithm can be successfully employed both for maximizing the throughput under a maximum transmitting power constraint [90] and for minimizing the power used while achieving a target transmission rate [85].

However, there exists two main problems with this approach. First, the proof convergence of this algorithm [79, 91] is subject to the assumption that the system operates in the weak interference regime; second, there exists sufficient literature [67, 92, 93] that shows that in DSCNs the operating point achieved though IWF is often inefficient.

1.1.2.7 Other Techniques

There are many other techniques that provide interesting opportunities for designing self configuring algorithms such as variational inequalities [94, 95], fuzzy logic [96, 97] are recently gaining growing attention. Furthermore, there exists a vast literature (e.g., [98, 99]), of heuristics techniques that can configure some particular parameters in some particular scenario. For instance in [43], a clustered multi-channel ad hoc network in which clusters are able to sense all available channels is considered. When an interference-free channel is not available, the choice on the channel is randomly

made. In low population density networks, this behavioral rule is shown to exhibit an acceptable performance with very little implementation complexity. Nonetheless, in high population-density networks, this approach is also shown to be highly suboptimal. However, their experience-based origin does not allow a theoretical analysis, hence their performance can be evaluated and compared only from a real world or simulation stand point. Moreover, the lack of theoretical understanding makes them unsuitable for configuring different kind of parameters.

1.1.2.8 Closing remarks

From this discussion, it is clear that there exist many viable options in order to design self configuring algorithms for DSCNs. Each approach differs from the other for, broadly, three characteristics: information assumptions, that is, what each algorithms needs to know on the environment, and what amount of information is necessary to exchange between the devices; the type of solution implemented with its respective performance; the possibility of theoretically analyzing the outcome of the algorithm. In order to establish which one of the previous approaches better fits a particular resource allocation problem would require a common theoretical background. This background could be exploited to compare the performance and balance the information assumptions and the computational onerousness of each algorithm. Unfortunately, a real comprehensive theoretical background is missing leaving simulations and prototyping as possible tests.

The approach followed in this thesis begins with the selection of GT as describing mathematical tool for the DSCN and adopts a learning algorithm as a framework to design a self-configuring algorithm.

The main motivations behind this are the following: (*i*) Compared to the other approaches GT and LT do not assume any particular configuration parameters, hence algorithms can be designed to set different parameters (e.g., channel, power, coding scheme); (*ii*) Equilibria analysis is an insightful tool in order to study the performance's limits of DSCNs; (*iii*) The other approaches often assume a certain level of coordination and information exchange between the different decision takers as a necessary condition.

1.2 Outline and contributions

This thesis is composed of five chapters: *Theory* (Chapter 2), *System Model* (Chapter 3), *Applications and Results* (Chapter 4) and *Conclusions and Outlook* (Chapter 5).

Chapter 2 introduces the theoretical background of the thesis. In Section 2.1, the game theoretical notations and concepts used throughout the thesis and a survey of

several important equilibrium concepts for DSCNs are introduced [7]. Section 2.2 divides learning algorithms into two groups: asymptotically converging and trial and error based. Moreover, it discusses several asymptotically converging learning algorithms and compares them in terms of information requirements and convergence properties. These algorithms' main limitations are identified in: The necessity of a particular structure of the underlying game in order to be able to predict the outcome and to insure convergence; The high level of information on the game each player must have in order for the algorithm to work.

Section 2.2.4 presents an algorithm that shows the feature of well performing in a vast variety of games, the trial and error (TE) learning algorithm. Among the foremost features of this algorithm, Theorem 2.10 [100, 101] shows that it is capable of stochastically converging to the NE that maximizes the performance of the network.

An abstract model of DSCNs is introduced and analyzed in Chapter 3. Section 3.1 provides the notations used throughout the whole thesis to describe a DSCN, and describes the optimization problem that defines the network's performance target. This optimization problem is given in a general form in order to be able to encompass several different possible goals, e.g, quality of service (QoS) provisioning with power consumption minimization, throughput maximization. For instance, this model can be used to represent both military and civil wireless networks. We assume that the goal of the designer is to maximize a certain global function that represents the quality of the communications in the network such as data rate or throughput, while minimizing the use of the resources, for instance the battery drain. The rationale behind this is that many real world applications require a minimum quality of the communication in order to function properly, for instance voice application and video application can require a minimum bit-rate, and their quality does not improve drastically once this minimum is exceeded. On the other hand, battery consumption is a key element in mobile wireless communication, and it is necessary to reduce the power drain in order to achieve long lasting communications. Clearly, this is of vital importance in military and emergency scenarios.

Moreover two instances of DSCN used as scenario to test the algorithms are detailed: a static dense DSCN and a mobile one. In Section 3.3 the game theoretical model of the DSCN is provided. A particular utility function (3.7) is specifically designed in order to have the NE with the highest social welfare [60] coinciding with the solution of the optimization problem expressed in Section 3.1. The goal is to exploit the TE learning algorithm's property of converging to the NE with the highest social welfare in order to converge to one of the solution of the optimization problem, thus configuring the network in an optimal way. A remarkable property of this utility function, is that, if each link in each cluster can evaluate its own QoS and transmit a 1 bit message to the cluster head (CH), then the CH is able to compute the value of the utility using only intra-cluster available information.

Chapter 4 presents the main results of the thesis. In Section 4.1 our theretical result regarding the TE algorithm are presented. Theorem 4.1 and Theorem 4.2 establish a precise link between the NE with highest social welfare and the solution of the optimization problem. This link allows a network designer to arbitrary select the goal of the network through the definition of the objective function and the constraints of the optimization problem discussed in Section 3.1. These two functions, in turn, define a utility function. This utility function will be used by TE to steer the network to the solution of the optimization problem.

Theroem 4.3 evaluates the upper and lower bound for the average number of iteration that the TE learning algorithm needs before reaching the NE for the first time, while Theorem 4.4 provides an approximation of the fraction of time the algorithm plays an NE. These two results, validated numerically in Section 4.1.3, are used to conclude that the two quantity depend on the experimentation parameter ε.

The algorithms presented in Section 2.2 are compared in terms of performance in Section 4.2. It i possible to observe that, in general, algorithms demanding higher level of information on the game's structure achieve more performing operating points. On the other hand, it is shown how the even algorithms that require high level of information on the game's structure can drop drastically in particular DSCNs. Section 4.3 tests the performance of TE with respect to the static and mobile DSCNs intrduced in Section 3.1. Its limits are identified in the instability of the channel-cluster association and in the non-optimal experimentation policy. In order to overcome these issues, an enhanced version of the algorithm is designed and thoroughly described in Section 4.5. This enhancment uses the insight gained from the theory developed in 4.1 to identify in the division of the experimentation probability a possible solution to the lack of stability of the algorithm solution. The enhancement is then tested and validated first against the standard TE learning algorithm in Section 4.5.3 then against other learning algorithms in Section 4.5.4, showing the algorithm's ability in configuring DSCNs.

The results are reported in Chapter 4. The TE learning algorithm is shown to be able to efficiently configure a DSCN. Some weaknesses due to the instability of the solution and a suboptimal policy of configuration selection are assessed. Therefore, a heuristic modification of the original algorithm is presented and its effectiveness in efficiently configuring a DSCN is shown. Furthermore, in Section 4.5.5 some preliminary results from a high fidelity simulator implemented in the context of the project CORASMA are reported validating the performance of the proposed solution on a realistic testbed. This thesis is finally concluded in Chapter 5 that summarizes the main results and provides an outlook to future work. In particular, the challenges with real systems are analyzed.

1.3 Publications

The work in this thesis is the result of several publications, and of the work conducted for the European defense agency (EDA) program CORASMA (COgnitive RAdio for dynamic Spectrum MAnagement). This work has triggered patents which are still under revision from the French ministry of defense. The main results of this thesis are summarized in the following articles.

1.3.0.9 Journal papers

- L. Rose, L., S. Lasaulce, S. M. Perlaza, M. Debbah, Learning equilibria with partial information in decentralized wireless networks, *IEEE Communications Magazine*, Vol 49, no. 8, pp. 136–142, Aug. 2011.
- L. Rose, L., S. M. Perlaza, C. J. Le Martret, M. Debbah, Self-Organization in Decentralized Networks: A Trial and Error Learning Approach *accepted for publication on IEEE Transactions on Wireless Communications*, 2013.

1.3.0.10 Conference papers

- L. Rose, L., S. M. Perlaza, M. Debbah, On the Nash equilibria in decentralized parallel interference channels, *Proc. of IEEE Workshop on Game Theory and Resource Allocation for 4G*, Kyoto, Japon, pp. 1–6, Jun. 2011.
- L. Rose, S. M. Perlaza, M. Debbah, C. J. Le Martret, Distributed power allocation with SINR constraints using trial and error learning, *in Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, pp. 1835–1840, Apr. 2012.
- L. Rose, C. J. Le Martret, M. Debbah, Channel and power allocation algorithms for *ad hoc* clustered networks, *in Proc. of the Military Communications and Information Systems Conference (MCC)*, Gdansk, Poland, pp. 1–8, 8–9 Oct. 2012.
- L. Rose, E. V. Belmega, W. Saad, M. Debbah, Dynamic service selection games in heterogeneous small cell networks with multiple providers, *in Proc. of the International Symposium on Wireless Communication Systems (ISWCS)*, Paris, France, pp. 1078–1082, 28–31 Aug. 2012.
- L. Rose, S. M. Perlaza, C. J. Le Martret, M. Debbah, Achieving Pareto optimal equilibria in energy efficient clustered *ad hoc* networks, *the IEEE International Conference on Communications Workshops (ICC)*, Budapest, Hungary, pp. 1491–1495, 1–8 Jun. 2013, .

- L. Rose, C. J. Le Martret, M. Debbah, Joint Channel and Power Allocation in Tactical Cognitive Networks: Enhanced Trial and Error, *in Proc. of the Military Communications and Information Systems Conference (MCC)*, Saint-Malo, France, pp. 1–11, 7–9 Oct. 2013, .

- L.Rose, R. Massin, L. Vijayandran, M. Debbah, C. J. Le Martret, CORASMA Program on Cognitive Radio for Tactical Networks: High Fidelity Simulator and First Results on Dynamic Frequency Allocation, *to appear in Proc. of the IEEE Military Communications Conference (Milcom)*, San Diego, CA, USA, 18–20 Nov. 2013

Chapter 2

Theory

This chapter introduces the theoretical background and notations used throughout the thesis. First, an introduction to GT is provided and different relevant equilibrium concepts are introduced. This includes the celebrated notion of NE, the correlated equilibria and the coarse correlated equilibria. Second, thanks to LT, a iterative processes converging to each of these equilibria are presented and analyzed. Third, the TE and optimal dynamic learning (ODL) learning algorithms are introduced and described and their main characteristics are explained. Given their characteristics of stochastically converging to a particular set of steady states, we design a particular utility function that allows these algorithms to steer the network to an efficient operating point. To this end, we provide analytical proofs of the ability of the TE learning algorithm to efficiently configure DSCNs.

2.1 Game Theory

2.1.1 Game Theory Introduction

GT is a mathematical framework that studies and provides analytical tools to predict the outcome of the complex interactions between rational autonomous entities known as players. The word rationality, here, demands the players to strictly adhere to a strategy based on perceived or measured results. In other words, players are decision-takers that aim at selfishly maximizing their own utility function, choosing their action among a set of possible choices called actions' set. Recently, GT has had a deep impact on various disciplines spanning from economics and engineering to sociology. The need to develop autonomous, distributed, and decentralized networks has given momentum to growing body of research, see among the others [14, 65, 102–107]. In general, GT can be divided into two branches: cooperative [61] and non-cooperative. In cooperative GT, players

can for coalitions cooperate in order to maximize their own utility functions, whereas, in non-cooperative, they must act independently.

Our main interests is to use GT as a tool for describing DSCNs in which the amount of message exchange between the decision takers is minimized. Therefore, non-cooperative GT fits the nature of our problem better.

2.1.2 Definitions and Notations

Hereunder, a brief review of some basic game-theoretical concepts used throughout the manuscript is provided.

There exist several possible representations of a game. The normal (or strategic) form is a convenient mathematical representation of a game defined as follows.

Definition 2.1 (Normal Form)**.** *A normal form game is defined by the triplet* $\mathcal{G} = \big(\mathcal{K}, \mathcal{A}, \{u_k\}_{k\in\mathcal{K}}\big)$*, where* $\mathcal{K}$ *is the set of players,* $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_K$ *is the actions' set, and* $\forall k \ \in \mathcal{K}$*,* $u_k : \mathcal{A}_k \to \mathbb{R}$ *is the utility function.*

We denote the vector of all actions as $\boldsymbol{a} = (a_1, a_2, \ldots, a_K)$, and we refer to it as action profile. In order to highlight the action taken by a particular player, with a slight abuse of notation, we use the notation $\boldsymbol{a} = (a_k, \boldsymbol{a}_{-k})$, where $\boldsymbol{a}_{-k}$ represents the vector containing the actions of all players except the k-th one.

An interdependent game (IG) is a game in which given an actions profile, any proper subset of players $\mathcal{K}^+$ can cause a utility change for some player that do not belong to $\mathcal{K}^+$ by a suitable change in their actions. In these games, in brief, there exists no group of players whose actions do not influence the utility of at least some other players in the game.

Definition 2.2 (Interdependent game)**.** $\mathcal{G}$ *is said to be interdependent if for every non-empty subset* $\mathcal{K}^+ \subset K$ *and every action profile* $\boldsymbol{a} = (\boldsymbol{a}_{K^+}, \boldsymbol{a}_{-K^+})$ *such that* $\boldsymbol{a}_{K^+}$ *is the action profile of all players in* $\mathcal{K}+$*, it holds that:*

$$\exists i \notin \mathcal{K}^+, \exists \boldsymbol{a}'_{K^+} \neq \boldsymbol{a}_{K^+} : u_i(\boldsymbol{a}'_{K^+}, \boldsymbol{a}_{-K^+}) \neq u_i(\boldsymbol{a}_{K^+}, \boldsymbol{a}_{-K^+}) \tag{2.1}$$

In non-cooperative games, each player $k \in \mathcal{K}$ selects its action $a_k \in \mathcal{A}_k$ in order to maximize its utility function $u_k(\boldsymbol{a})$ in an independent selfish manner. The global performance of an action profile is measured through the social welfare function $W : \mathcal{A} \to \mathbb{R}$, defined as follows.

Definition 2.3. For any action profile $\boldsymbol{a} \in \mathcal{A}$ its social welfare is defined as:

$$W(\boldsymbol{a}) = \sum_{k\in\mathcal{K}} u_k(\boldsymbol{a}). \tag{2.2}$$

For a network designer, where the action profile represents a particular configuration in a DSCN and the utility function the performance of the communications, actions associated with higher social welfare values are generally more appealing. For instance, in a DSCN , the set of players could consist of the set of wireless terminals present in the network, the action set could be any feasible vector of transmit powers, and the utility function could be the spectral efficiency. Other components are also possible and they depend on the scope and purpose of the network design.

Let $\triangle(\mathcal{A})$ denote the set of all possible probability distributions over the whole set of actions $\mathcal{A}$, and $\triangle(\mathcal{A}_k)$ represents the set of all possible probability distributions of user k over its action set. The elements of the set $\mathcal{A}_k$ are referred to as the *actions* of player k and those of the set $\triangle(\mathcal{A}_k)$ as the *strategies* of player k. A given strategy of player k is denoted by $\boldsymbol{\pi_k} = (\pi_{k,A_k^{(1)}}, ..., \pi_{k,A_k^{(N_k)}}) \in \triangle(\mathcal{A}_k)$, where $\pi_{k,A_k^{(n_k)}}$ represents the probability that player k plays action $A_K^{(n_k)}$. Indicate by $\boldsymbol{\phi} = (\phi_{A^{(1)}}, ..., \phi_{A^{(N)}}) \in \triangle(\mathcal{A})$, with $N = \prod_{j=1}^{K} N_j$, a given joint probability distribution over the set $\mathcal{A}$, with $\phi_{A^{(n)}}$ being the probability of observing $A^{(n)}$ as an outcome of the game.

The most general type of equilibria used in this thesis is the coarse correlated equilibrium (CCE) [77]. The idea behind CCE is that actions chosen by the players of a game may be statistically correlated. For instance, correlation may appear when a common broadcast signal is observed by several transmitters choosing their transmit configuration, e.g., a power control policy. The signals received by the players are referred to as recommendations. In such a context, a CCE is a probability distribution $\boldsymbol{\phi} \in \triangle(\mathcal{A})$ over the set of action profiles of the game from which no player has interest in unilaterally deviating. The realizations of this joint distribution $\boldsymbol{\phi}$ are the recommendations. Mathematically, this can be written as follows.

Definition 2.4 (Coarse Correlated Equilibrium)**.** *A joint probability distribution* $\boldsymbol{\phi} \in \triangle(\mathcal{A})$ *is a CCE if* $\forall k \in \mathcal{K}$ *and* $\forall a'_k \in \mathcal{A}_k$ *it holds that*

$$\sum_{\boldsymbol{a} \in \mathcal{A}} u_k(\boldsymbol{a}) \phi_{\boldsymbol{a}} \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a}_{-k}) \phi_{-k,\boldsymbol{a}_{-k}}, \tag{2.3}$$

where $\phi_{-k,a_{-k}} = \sum_{a_k \in \mathcal{A}_k} \boldsymbol{\phi}_{(a_k,\boldsymbol{a}_{-k})}$ *is the marginal probability distribution with respect to* a_k*.*

Following the notion of CCE, players are assumed to decide, *before* receiving the recommendation, whether to commit to follow it or not. At a CCE , all players are willing to commit to follow the recommendation given that all the others also choose to commit. That is, if a single player decides not to commit to follow the recommendations, it experiences a lower (expected) utility.

A special case of CCE is the correlated equilibrium (CE) , [77]. The difference between the CCE and the CE is that, in the latter, players choose whether to follow or

not a given recommendation, *after* it has been received. Therefore, there is no *a priori* commitment. It follows in particular that, for a given game, the set of all CE is a subset of the set of all CCE [77].

Definition 2.5 (Correlated Equilibrium)**.** *A joint probability distribution $\boldsymbol{\phi} \in \triangle(\mathcal{A})$ is a CE if $\forall k \in \mathcal{K}$ and $\forall a'_k, a_k \in \mathcal{A}_k$ it holds that*

$$\sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a_k, \boldsymbol{a_{-k}}) \phi_{a_k, \boldsymbol{a_{-k}}} \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \phi_{a_k, \boldsymbol{a_{-k}}}. \tag{2.4}$$

If the players choose their strategy following independent individual probability distributions $\boldsymbol{\pi}_k \in \triangle(\mathcal{A}_k)$, i.e., $\phi_a = \prod_{j=1}^{K} \pi_{j,a_j}$ in (2.3), we obtain from Definition 2.4, the definition of mixed Nash equilibrium (MNE) [62] or Nash equilibrium in mixed strategy. The MNE is a special case of CE, hence a special case of CCE. In detail, a MNE is a vector of individual probability distributions $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K)$ which is stable to unilateral deviations. This means that if any player k adopts a different probability distribution from the corresponding $\boldsymbol{\pi}_k$, then it observes a lower (expected) utility.

Definition 2.6 (Nash Equilibrium in mixed strategy)**.** *A vector of probability distributions $\boldsymbol{\pi} = (\pi_{1,a_1}, \ldots, \pi_{K,a_K})$ is a MNE if $\forall k \in \mathcal{K}$ and $\forall a'_k, a_k \in \mathcal{A}_k$ it holds that*

$$\sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j} \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j}. \tag{2.5}$$

Definition 2.6 is non-orthodox definition of the MNE. A proof of the equivalence between Definition 2.6 and a more standard formulation is provided in Appendix A.

As shown in [74], this type of equilibria always exists in games with finite number of players and finite action sets. For more results on the existence and multiplicity of MNE, the reader is referred to [108]. The finiteness assumption is especially relevant when a wireless terminal has to select a given communication setting, e.g., a logical channel, a constellation size, or a transmit power level[1].

A refinement of the concept of MNE is the $\epsilon-$*equilibrium*. An $\epsilon-$equilibrium is a mixed strategy profile $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K) \in \triangle(\mathcal{A}_1) \times \ldots \times \triangle(\mathcal{A}_K)$ such that if only one player k uses a different strategy from its corresponding $\boldsymbol{\pi}_k$, it does not observe a utility improvement greater than $\epsilon > 0$. An instance of $\epsilon-$NE is the logit equilibrium [77].

Definition 2.7 (ϵ-Equilibrium)**.** *A vector of probability distributions $\boldsymbol{\pi} = (\pi_{1,a_1}, \ldots, \pi_{K,a_K})$ is an ϵ-Equilibrium if $\forall k \in \mathcal{K}$ and $\forall a'_k, a_k \in \mathcal{A}_k$ it holds that*

$$\sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j} \geq \sum_{a_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j} + \epsilon. \tag{2.6}$$

[1] In real communications the transmit power is always expressed by a finite number of bits, hence the power levels can be taken from a finite set.

The concept of NE is obtained by restricting the players to deterministically choose one of their actions instead of choosing it by following a probability distribution. A NE is therefore a special case of MNE where the individual probability distribution is a Dirac's delta function over a given action. Therefore, a NE is a vector of actions $\boldsymbol{a^*} = (a_1^*, \ldots, a_K^*)$ stable to unilateral deviations, i.e., if any player k adopts a different action from its corresponding a_k^*, while the others keep their equilibrium actions, player k observes a lower (instantaneous) utility Its definition is given hereunder.

Definition 2.8 (Nash Equilibrium in pure strategy)**.** *An action profile $\boldsymbol{a^*} \in \mathcal{A}$ is a NE if $\forall k \in \mathcal{K}$ and $\forall a'_k, a_k \in \mathcal{A}_k$ it holds that*

$$u_k(a_k, \boldsymbol{a_{-k}}) \geq u_k(a'_k, \boldsymbol{a_{-k}}). \tag{2.7}$$

A summary of the equilibrium concepts introduced can be found in Figure 2.1. Since the NE is a strategy profile such that no player can improve its utility by a unilateral deviation, it represents an operating point that is both predictable and stable. This means that, once the system achieves the NE , there exists no user that has any incentive to deviate from the action profile, thus the system state does not evolve any further.

However, in general, the NE performance is suboptimal compared with the performance of a theoretical optimum. More desirable action profiles that cope with these issues are the Pareto optimal states. An action profile is said to be Pareto optimal if it is not possible to increase the utility of a player without decreasing the utility of another.

Definition 2.9 (Pareto optimality)**.** *An action profile $\boldsymbol{a^{(1)}} \in \mathcal{A}$ is Pareto optimal if it does not exist $\boldsymbol{a^{(2)}} \in \mathcal{A}$ such that $\forall\ k \in \mathcal{K}, a^{(2)} \geq a^{(1)}$.*

Unfortunately, Pareto optimal action profiles are not necessarily stable. In fact, unless it also a NE, the player that can improve its utility function will do it at the expense of the other players leasing to a non-Pareto state. Remarkably, in some cases, in order to improve the performance of the NE of a game it is sufficient modify the game by reducing the dimension of the action space [93, 109, 110]. In general, this is done by eliminating the most inefficient NE.

2.2 Learning Theory

2.2.1 Learning Theory Introduction

As highlighted in Section 2.1, computing equilibria for non-cooperative games requires both rationality and full knowledge on the structure of the game from the players. In

practical terms, in a DSCN, this means devices that are perfectly aware of the performance of any possible configuration. Hence, iterative procedures that require little or no prior information on the game and that may converge to a predictable equilibrium are an appealing solution. Hereunder, we discuss some of the basic learning algorithms present in the literature. We divide the learning algorithms into two types: the asymptotic learning algorithms and the trial and error based ones. The main difference relies on the fact that the formers converge to a steady state *asymptotically*. This means that the learning process is divided into two distinguishable phases: an exploring phase, in which the algorithms try to learn the equilibrium and an exploiting phase during which the achieved equilibrium is used as a configuration. TE, on the other hand, follows a different philosophy. The exploitation of the configuration is done at run time. Here, the equilibrium is not achieved in the long run, rather the equilibrium is played with high probability a large portion of the time [111].

2.2.2 Asymptotic Learning Algorithms

The process of learning equilibria is basically an iterative process. Each iteration of the learning process can be broadly divided into three phases: (i) the observation of the environment at iteration t, which evaluatesthe performance of the action chosen at time $t-1$; (ii) the improvement of the strategy $\boldsymbol{\pi}_k(t)$ based on the current observation and (iii) the selection of the action $a_k(t)$ according to the strategy $\boldsymbol{\pi}_k(t)$. Hence, we say that players learn to play an equilibrium, if after a given number of iterations, the strategy profile $\boldsymbol{\pi}(t) = (\boldsymbol{\pi}_1(t), \ldots, \boldsymbol{\pi}_K(t)) \in \triangle(\mathcal{A}_1) \times \ldots \times \triangle(\mathcal{A}_K)$ converges to an equilibrium strategy.

The purpose of this section is to introduce the following set of learning algorithms: BRD, fictitious play (FP), smoothed fictitious play (SFP), regret matching (RM), reinforcement learning (RL) and the joint utility and strategy estimation reinforcement learning (JUSTE-RL). In Section 4.2, we compare such algorithms in terms of relevant features in the context of wireless communications. For instance, type of observations, type of action sets, convergence time, nature of the steady state achieved when convergence is observed and conditions for convergence.

2.2.2.1 Best Response Dynamics

In its most basic form, the BRD [112] relies on the following assumptions: at each game stage $t \in \mathbb{N}$, every player k plays the action $a_k(t)$ which optimizes its own utility function given the actions played by the other players. When all players play simultaneously at each stage (simultaneous-BRD), the optimization of player k is done with respect to the action profile $\boldsymbol{a}_{-k}(t-1)$. When players play sequentially, only one player at each stage (sequential-BRD) updates its action $a_k(t)$, optimizing it with respect to the action

profile $(a_1(t), \ldots, a_{k-1}(t), a_{k+1}(t-1), \ldots, a_K(t-1))$. Note that observing the actions of the other players is not always necessary. In some cases [113], only an aggregate function of all the other players' actions is needed to implement the BRD. As a relevant case, the IWF can be considered a particular case of the BRD , when the utility function is the spectral efficiency and the action set is composed of the power profile over the available channels.

2.2.2.2 Fictitious Play

The FP [114] is an iterative procedure in which each player *believes* that all the other players play following a fixed probability distribution. As a consequence, each player's goal is to estimate this probability distribution, thus learning its own optimal optimal action. This algorithm relies on the assumptions that at each stage t, each player k knows all the past actions of all the other players, i.e., $a_j(0), \ldots, a_j(t-1), \forall\, j \in \mathcal{K} \setminus \{k\}$. Based on such observations, player k calculates the empirical frequencies with which each player plays its corresponding actions. These empirical frequencies are referred to as beliefs. Let us denote the belief that player $k \neq j$ has on the probability distribution of player j by the vector $\boldsymbol{f}_j(t) = \left(f_{j,A_j^{(1)}}(t), \ldots, f_{j,A_j^{(N_j)}}(t) \right) \in \triangle(\mathcal{A}_j)$. At each stage, all players (simultaneously or sequentially, as in the BRD) choose their current action by optimizing their expected utility with respect to the beliefs on all the other players, i.e., $a_k(t) \in \arg\max_{a_k \in \mathcal{A}_k} \mathbb{E}_{\boldsymbol{f}(t)} \left[u_k \left(a_k, \boldsymbol{a}_{-k} \right) \right]$, where $\boldsymbol{f}(t) = (\boldsymbol{f}_1(t), \ldots, \boldsymbol{f}_K(t))$.

2.2.2.3 Smooth Fictitious Play

The convergence of FP is not ensured in games with cycles and its ability to explore the whole action set is highly constrained [7, 77, 115]. To overcome these issues, a simple variation of the FP has been proposed under the name of SFP. The assumptions on which SFP relies on are the same as FP and actions can be updated either simultaneously or sequentially. The main difference between SFP and FP is that, at each stage t, player k does not choose a deterministic action. It rather builds a probability distribution $\boldsymbol{\pi}_k(t) \in \triangle(\mathcal{A}_k)$ to choose its action $a_k(t)$. Such a probability distribution can be interpreted as the one that maximizes a weighted sum of the original expected utility and other continuous strictly concave function. For instance, if such a function is the *entropy function* [77], the resulting probability distribution is given by the logit probability distribution.

2.2.2.4 Regret Matching

Contrary to the case of BRD , FP and SFP, where players determine whether to play or not a particular action based on the idea of utility maximization, in RM [116], such a

decision is made considering the notion of regret minimization. The regret that player k associates with action $A_k^{(t_k)}$ is defined as the difference between the average utility the player would have obtained by always playing $A_k^{(t_k)}$ and the average utility actually achieved with the current strategy, i.e.,

$$r_{k,A_k^{(n_k)}}(t) = \frac{1}{n-1}\sum_{t=1}^{n-1}(u_k(A_k^{(n_k)}, \boldsymbol{a}_{-k}(t)) - u_k(a_k(t), \boldsymbol{a}_{-k}(t))). \tag{2.8}$$

This algorithm relies on the assumptions that, at every stage t, player k is able to evaluate its own utility, i.e., to calculate $u_k(a_k(t), \boldsymbol{a}_{-k}(t))$, and to compute the utility it would have obtained if it had played any other action a'_k, i.e. $u_k(a'_k, \boldsymbol{a}_{-k}(t))$. Finally, the action to be played at stage t is taken following the probability distribution $\boldsymbol{\pi}_k(t)$, which is obtained by normalizing to one the regret vector $\boldsymbol{r}_k(t) = \left(r_{k,A_k^{(1)}}(t), \ldots, r_{k,A_k^{(N_k)}}(t)\right)$. Even though regret minimization is an appealing characteristic, even no-regret points need not to reflect optimal operating conditions for multi-agent systems [117].

2.2.2.5 Reinforcement Learning

In the case of RL [80, 118], players are modeled as automata that implement a given behavioral rule without any rationality. In general, RL techniques rely on the following two conditions: (i) for each player k, the action set $\mathcal{A}_k$ is finite and for all action profiles $\boldsymbol{a} \in \mathcal{A}$, the achieved utility $u_k\left(a_k, \boldsymbol{a}_{-k}\right)$ is bounded; (ii) each player is able to periodically observe its own achieved utility. Intuitively, the idea behind RL is that actions leading to higher utility observations in stage t are granted with higher probabilities in the game stage $t+1$, and vice versa.

2.2.2.6 Joint Utility and Strategy Estimation - Reinforcement Learning

A variant of the RL algorithm, JUSTE-RL [119] relies on the same assumptions as the classical RL. The main difference between classical RL and JUSTE-RL is that, in the former, the observation $\tilde{u}_k(t)$ of the utility of player k is used to directly modify the probability distribution $\boldsymbol{\pi}_k(t)$; in the latter, such an observation is used to build an estimation of the expected utility for each of the actions. Such utility estimates are then used in the same iteration to finally build a probability distribution $\boldsymbol{\pi}_k(t)$ from which action $a_k(t)$ will be drawn. Thus, each player always possesses an estimation of the expected utility it obtains by playing each of its actions.

2.2.3 Discussion

The purpose of this section is to provide additional insights about the performance and pertinence of the learning algorithms described above in the context of decentralized

wireless networks. In the following, we compare the algorithms in terms of several fundamental features. We summarize this discussion in Table 2.1.

2.2.3.1 Observations

At each iteration of a given learning algorithm, each player must obtain some information about how the other players are reacting to its current action, in order to update their strategy and choose the following action. Broadly speaking, in algorithms such as BRD, FP, SFP and RM, players must, in general, observe the actions played by all the other players. This implies that a large amount of additional signaling is required to broadcast such information in wireless networks, or that sensing information must be precise and reliable. In some particular cases, this condition can be relaxed and less information is required [65, 83]. However, this is highly dependent on the topology of the network and the explicit form of the utility function [79]. Other algorithms, such as RL and JUSTE-RL , only require that each player observes its corresponding achieved utility at each iteration. This is in fact, their main advantage, since such information requires a simple feedback message from the receiver to the corresponding transmitters [80, 119].

2.2.3.2 Knowledge and Calculation Capabilities

Learning algorithms such as BRD, FP, SFP and RM involve an optimization problem at each iteration [112]. This means at each algorithm's iteration the players need to compute either the maximization of the (expected or instantaneous) utility or minimization of the regret. Therefore, in general, highly demanding calculation capabilities are required to implement them. More importantly, solving such optimization requires the knowledge of the closed-form expression of the utility function. This implies that, in general, each player must be provided with knowledge on the structure of the game, i.e., set of players, action sets, current strategies, channel realizations, etc. In this respect, RL and JUSTE-RL algorithms are more attractive since only algebraic operations are required to update the strategies. In terms of knowledge, in both RL and JUSTE-RL, players are only required to know the action they actually played at the previous iteration and the corresponding achieved utility. Indeed, it is possible to say that players are not even aware of the presence of other players.

2.2.3.3 Nature of the Action Sets

The nature of the action sets of the game plays an important role. The BRD can be used for both continuous and discrete action sets, whereas in their standard versions FP, SFP, RM, RL, and JUSTE-RL are designed for discrete action sets. For instance, action sets are discrete in problems where a channel, constellation size or discrete power

levels must be selected, whereas continuous sets are more common in power allocation problems [108].

2.2.3.4 Steady State

When a steady state is achieved by one of the algorithms under consideration, such state may correspond to one of the equilibrium notions presented in Section 2.1.2. In particular, when BRD and FP converge, the strategy of the players at the steady state is a NE [120]. In the case of the RM, it converges to an element of the set of CCE [77]. Relevantly, even though the notion of CCE relies on the idea of the recommendations studied in Section 2.1.2, this algorithm does not require the existence of recommendations to converge to a CCE. When SFP or JUSTE-RL achieve a steady state, it corresponds to an ε-NE [112]. On the contrary, in the case of RL, a steady state not necessarily corresponds to a particular notion of equilibrium [118]. A summary of the steady states of the algorithms is represented in Figure 2.1.

2.2.3.5 Convergence Conditions

Regarding the conditions for convergence, only sufficient conditions are available. As shown in Table 2.1, the considered algorithms typically converge in certain classes of games [77] such as dominant-solvable gamess (DSGs), potential gamess (PGs), super modular gamess (SMGs), zero sum gamess (ZSGs) [77].

2.2.3.6 Synchronization

In the particular case of algorithms where each player must observe the actions of the others, e.g., BRD, FP, SFP and RM, certain synchronization is required in order to allow players to know when to play and when to observe the actions of the others. In wireless communications, this requirement implies the existence of a given protocol for signaling messages exchange. Conversely, when players require only an observation of their individual utility, such synchronization between all the players becomes irrelevant. Here, only a feedback message from the receiver to the corresponding transmitters per learning iteration is sufficient.

2.2.3.7 Environment

Learning techniques such as the BRD are highly constrained for real system implementations since they require the network to be static during the whole learning processes. On the contrary, all the other techniques allow the dynamics of the network to be captured by their statistics as long as they are stationary. This is basically because, contrary

	BRD	FP	SFP
Observations	$\boldsymbol{a}_{-k}(t)$	$\boldsymbol{a}_{-k}(t)$	$\boldsymbol{a}_{-k}(t)$
Closed Expression for u_k	Yes	Yes	Yes
Computation complexity	Optimization	Optimization	Optimization
Steady State	NE	NE	ε-NE
Condition for Convergence	DSG, PG, SMG	DSG, PG, ZSG	DSG, PG, ZSG
Synchronization to Play	Yes	Yes	Yes
Environment	Static	Stationary	Stationary
	RM	RL	JUSTE-RL
Observation	$\boldsymbol{a}_{-k}(t)$	$\tilde{u}_k(t)$	$\tilde{u}_k(t)$
Closed Expression for u_k	Yes	No	No
Computation complexity	Optimization	Algebraic Operation	Algebraic Operation
Steady State	CCE	$--$	ε-NE
Condition for Convergence	$--$	$--$	DSG, 2$-$player ZSG, PG
Synchronization to Play	Yes	No	No
Environment	Stationary	Stationary	Stationary

TABLE 2.1: Benchmark of Asymptotic Learning Algorithms.

to BRD, all the other techniques determine whether to play or not a particular action based on the expected utility rather than the instantaneous utility.

2.2.3.8 Convergence Speed

The speed of convergence (when it is observed) is highly influenced by the amount of information available for the players. For instance, FP, SFP and RM converge faster than JUSTE-RL since the formers calculate the expected utility relaying on a closed form expression. Conversely, the latter calculates it as the time-average of the instantaneous observations of the achieved utility. This requires a large number of observations to obtain a reliable approximation of the expected utility. We do not state any particular comment on the speed of convergence of BRD and RL since, in the former, the scenario is considered fixed and in the latter, it does not necessarily converge to an equilibrium strategy.

2.2.4 State machine based algorithms

The purpose of this section is twofold. First it provides a brief description of two algorithms based on a particular state machine, the TE learning algorithm and the ODL algorithm. Second it provides some basic theoretical results justifying their use in DSCNs.

2.2.4.1 Trial and Error Description

The TE learning algorithm can be described by a state machine locally implemented by each player. The main feature of this state machine is that the set of stochastically stable states are the NE that maximize the social welfare.

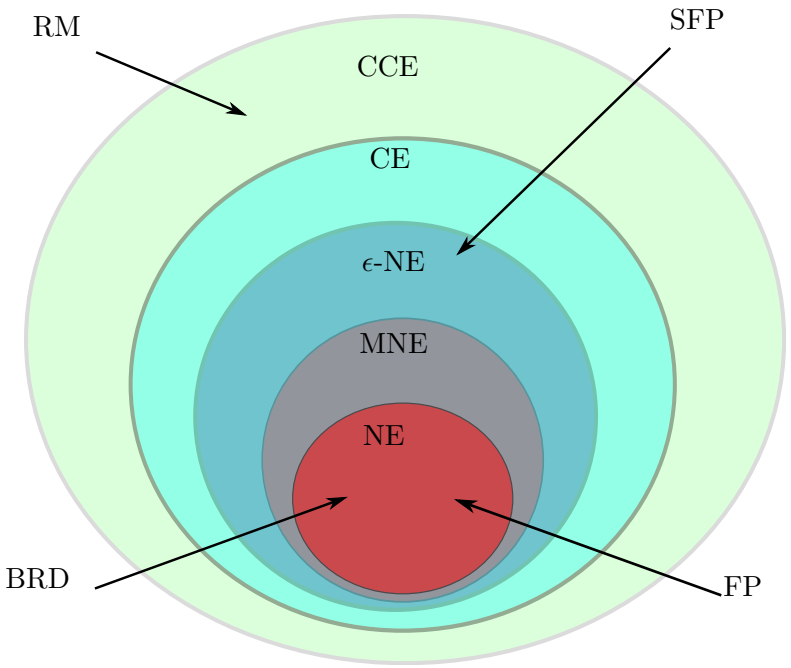

FIGURE 2.1: Summary of types of equilibria and relative asymptotic learning algorithms.

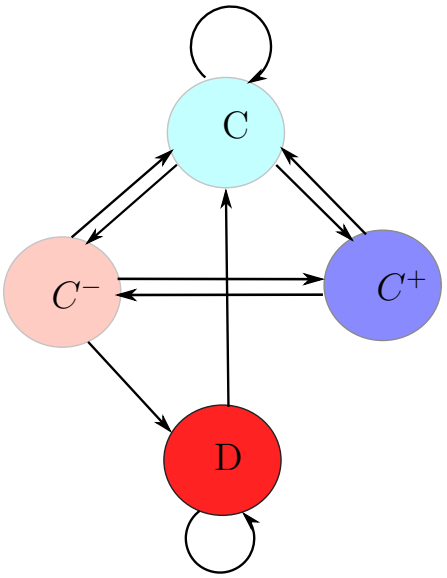

FIGURE 2.2: TE learning algorithm possible transitions.

At each iteration t, the state of player k is defined by the triplet:

$$Z_k(t) = \{m_k(t), \bar{a}_k(t), \bar{u}_k(t)\}\,, \tag{2.9}$$

where $m_k(t) \in \{C, C^+, C^-, D\}$ represents the *mood*: *content* (C), *hopeful* (C^+), *watchful* (C^-), *discontent* (D), $\bar{a}_k(t) \in \mathcal{A}$ and $\bar{u}_k(t) \in [0, 1]$ represent the *benchmark* action and *benchmark* utility, respectively. The state machine transitions and behavior are detailed hereunder and the possible transitions are summarized in Figure 2.2. Note that the notation $a \Leftarrow b$ indicates that variable a takes the value of variable b.

Content: Let $\varepsilon \in [0,1]$ be an experimentation parameter and assume that the state of player k at time $t-1$ is $Z_k(t-1) = \{C, \bar{a}_k(t-1), \bar{u}_k(t-1)\}$. Then, at iteration t, it selects its action according to the following rule: with probability $(1-\varepsilon)$, it plays the benchmarked action $a_k(t) = \bar{a}_k(t-1)$ or with probability ε, it plays another randomly selected action $a_k(t) \neq \bar{a}_k(t-1)$. Once player k has played action $a_k(t)$, it observes the value of its utility function $u_k(t)$.

The player updates its state as follows: If $a_k(t) \neq \bar{a}_k(t-1)$ and $u_k(t) \leq \bar{u}_k(t-1)$, then $Z_k(t) \Leftarrow Z_k(t-1)$; If $a_k(t) \neq \bar{a}_k(t-1)$ and $u_k(t) > \bar{u}_k(t-1)$, then, with probability $\varepsilon^{G(u_k(t)-\bar{u}_k(t-1))}$, it sets $m_k(t) \Leftarrow m_k(t-1)$, $\bar{a}_k(t) \Leftarrow a_k(t)$ and $\bar{u}_k(t) \Leftarrow u_k(t)$, while with probability $\left(1-\varepsilon^{G(u'_k(t)-\bar{u}_k(t-1))}\right)$, it sets $Z_k(t) \Leftarrow Z_k(t-1)$; If $a_k(t) = \bar{a}_k(t-1)$ and $u_k(t) \geq \bar{u}_k(t-1)$ then, $m_k(t) \Leftarrow C^+$, $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$, $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$; If $a_k(t) = \bar{a}_k(t-1)$ and $u_k(t) < \bar{u}_k(t-1)$ then $m_k(t) \Leftarrow C^-$, $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$, $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$.

Note that if player k does not experiment (it plays its benchmarked action) and its utility increases, then it becomes *hopeful*, while if it decreases, it becomes *watchful*. Here, the function $G : \mathbb{R} \to \mathbb{R}$ must be such that:

$$0 \leq G(x) < \frac{1}{2}. \tag{2.10}$$

Numerical simulations suggest that a linear formulation such as: $G(\Delta u) = -0.2\Delta u + 0.2$, with $\Delta u = u_k(t) - \bar{u}_k(t-1)$, performs well under several scenarios.

Hopeful: Assume that the state of player k at time $t-1$ is described by the following triplet: $Z_k(t-1) = \{C^+, \bar{a}_k(t-1), \bar{u}_k(t-1)\}$. Then, at iteration t, it plays the benchmark action $a_k(t) = \bar{a}_k(t-1)$ and it observes the value of its utility function $u_k(t)$. If $u_k(t) \geq \bar{u}_k(t-1)$ then, $m_k(t) \Leftarrow C$, $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$ and $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$; otherwise, $m_k(t) \Leftarrow C^-$, $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$ and $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$.

Watchful: Assume that the state of player k at time $t-1$ is described by the following triplet: $Z_k(t-1) = \{C^-, \bar{a}_k(t-1), \bar{u}_k(t-1)\}$. Then, at iteration t, it plays the benchmark action $a_k(t) = \bar{a}_k(t-1)$ and it observes the value of its utility function $u_k(t)$. If $u_k(t) > \bar{u}_k(t-1)$, then $m_k(t) \Leftarrow C^+$, $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$ and $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$; otherwise, $m_k(t) \Leftarrow D$, $\bar{u}_k(t) \Leftarrow \bar{u}_k(t-1)$ and $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$.

Discontent: Assume that the state of player k at time $t-1$ is described by the following triplet: $Z_k(t-1) = \{D, \bar{a}_k(t-1), \bar{u}_k(t-1)\}$. Then, at iteration t, it randomly selects an action $a_k(t)$ and observes the value of its utility function $u_k(t)$. The state is updated as follows: with probability $p = \varepsilon^{F(u_k(t))}$ it sets $m_k(t) \Leftarrow C$, $\bar{u}_k(t) \Leftarrow u_k(t)$ and $\bar{a}_k(t) \Leftarrow \bar{a}_k(t-1)$; with probability $(1-p)$ it sets $m_k(t) \Leftarrow D$, $\bar{u}_k(t) \Leftarrow u_k(t)$ and $\bar{a}_k(t) \Leftarrow a_k(t)$. The function $F : \mathbb{R} \to \mathbb{R}$ must be such that

$$0 \leq F(u) < \frac{1}{2K}. \tag{2.11}$$

Numerical simulations suggest that a linear formulation such as: $F(u) = -\frac{0.2}{K}u + \frac{0.2}{K}$ performs well under several scenarios.

2.2.4.2 Convergence of the Trial and Error Learning Algorithm

This section discusses the properties convergence points of the TE learning algorithm. In [111] and [101], the authors proved that the stochastically stable action profiles of the trial and error algorithm (i.e., action profiles that are played with high probability most of the time) are those NE that maximize the social welfare. Theorem 2.10 restates their main results.

Theorem 2.10. *Let the interdependent game $\mathcal{G}$ have at least one pure NE and let each player use TE. Then, for each ε small enough, there exists a δ such that a pure Nash equilibrium that maximizes the sum utility among all equilibrium states is played $(1-\delta)$ fraction of the time.*

Theorem 2.10 states that if all players implement the TE algorithm and there exists at least one NE, then the NE with the highest social welfare is played during a *large* fraction of the time. In general, the quantity $1-\delta$ depends on ε and on the particular game $\mathcal{G}$. When players implement the TE algorithm, the notion of convergence largely differs from the classical idea of convergence, that is, a dynamic distance minimization with respect to certain action profile (e.g., an NE, a correlated equilibria, etc). With those algorithms, once the steady state is reached, the action profile remains the same. The convergence of the TE algorithm must be understood in terms of the time players remain at a given action profile. Indeed, the system can be at an NE, but it might arbitrarily leave it to experiment other action profiles. Therefore, in this setting, convergence refers to the fact that the system remains on certain action profiles a large fraction of the time.

This seemingly non-appealing feature turns out to be a strong point of the procedure if one considers that a wireless system in general, and a DSCN in particular, is by definition a non-stationary system. This means that equilibria and their performance tend to change with time due to unpredictable factors. Algorithms that tend to be static once a steady state is reached may therefore force the network to use a strongly sub-optimal configuration until the learning procedure is reinitialized. On the contrary, an algorithm that keeps learning and updating its working point continuously could react more quickly to the change of the communication conditions.

2.2.5 Optimal Dynamic Learning

For DSCNs in which resources are very scarce, for instance a network with a great imbalance between users and channels available, an algorithm that aims at implementing

an equilibrium might not be efficient due to the global performance limitations. As a consequence, this section presents an algorithm whose stochastically stable points are all the Pareto optimal action profiles.

2.2.5.1 Optimal Dynamic Learning description

In ODL, every player k implements a state machine, where a state $Z_k(t) = (m_k(t), a_k(t), u_k(t))$ is defined by a triplet composed by a *mood* $m_k(t)$, a *benchmark* utility $\bar{u}_k(t)$ and a *benchmark* action $\bar{a}_k(t)$. Transitions between the states happen when a change occurs in the utility as a consequence of a variation in the network (e.g., fading, a player switches its channel). There are two possible moods: content (C) and discontent (D).

Content: If at time t player k is content, it chooses action $a_k(t)$ following the probability distribution

$$\pi_{k,a_k} = \begin{cases} \frac{\epsilon^{K+1}}{|\mathcal{A}_k|-1} & \text{if} \quad \bar{a}_k \neq a_k \\ 1-\epsilon^{K+1} & \text{if} \quad \bar{a}_k = a_k. \end{cases}, \tag{2.12}$$

where $\pi_{k,a_k} = \Pr(a_k(t) = \bar{a}_k(t))$. In the case in which $\bar{a}_k(t) = a_k(t)$ and $\bar{u}_k(t+1) = u_k(t+1)$ (i.e., it did not experiment and the utility has not changed), then $m_k(t+1) = C$, $\bar{a}_k(t+1) = \bar{a}_k(t)$ $\bar{u}_k(t+1) = \bar{u}_k(t)$. Otherwise, if $\bar{a}_k(t) \neq a_k(t)$ or $\bar{u}_k(t+1) \neq u_k(t+1)$, the player updates the benchmark utility and action with the new values, then it remains *content* with probability $\epsilon^{(1-u_k(t))}$ or it becomes *discontent* with probability $1-\epsilon^{(1-u_k(t))}$.

Discontent: If at time t player k is discontent, it chooses action $a_k(t)$ with uniform probability among all its possible choices. Then, with probability $\epsilon^{(1-u_k(t+1))}$ the mood changes to content, and $a_k(t)$ and $u_k(t+1)$ become the new benchmark action and utility, while, with probability $1-\epsilon^{(1-u_k(t+1))}$, the mood remains discontent.

2.2.5.2 Optimal Dynamic Learning Convergence

The algorithm previously described shows some useful properties shown in [121]; for the sake of simplicity, we rewrite the main result within with our notation.

Theorem 2.11. Let $\mathcal{G}$ be an interdependent K-person game on a finite joint action space $\mathcal{A}$. Under the dynamics defined by ODL, a state Z is stochastically stable if and only if the following conditions are satisfied:
(i) The action profile $\boldsymbol{a}$ maximizes $W(\boldsymbol{a}) = \sum_{k\in\mathcal{K}} u_k(\boldsymbol{a})$
(ii) The mood of each agent is content, i.e., $m_k = C$ $\forall k \in \mathcal{K}$.

The concept of stochastic stability, introduced in [100], is at the base of the algorithm. Broadly, a stochastically stable action profile is an action profile that, once it is

reached by the algorithm, there is a small probability of leaving it. Note that, compared with other results in the literature, for instance [85], [67], this algorithm does not focus on reaching a NE. Thus the action profiles most implemented by ODL have, generally, a higher social welfare than those implemented by NE-focused algorithms. On the other hand, social welfare maximizing action profiles, generally, are not individually optimum, thus they are intrinsically less stable than NE.

2.3 Closing Remarks

In this chapter, basic game theoretical definitions have been introduced as well as several notions of equilibrium and different iterative procedures known as learning algorithms. The iterative repetition of these algorithms allows the games' players to achieve such equilibria with minimal knowledge on the game structure. In particular, a most general notion of equilibrium, namely, the CCE was introduced. Therefore, the CE was discussed as particular case of the coarse correlated equilibrium. The MNE, the ϵ-equilibrium and the NE were also defined and characterized.

The learning algorithms have been divided in two different groups, asymptotic learning algorithms and state machine based learning algorithms. While the algorithms of the first group (namely the BRD, the FP, the SFP, RL and JUSTE-RL) achieve their steady state in the long run, the algorithms of the second group guarantee that a steady state is played with high probability. Moreover, we have presented a theorem that grantees that the most probable action profile played by TE is the NE that maximize the social welfare. The pertinence of these algorithms for DSCN has been identified in terms of system constraints (continuous or discrete actions, required information, synchronization, signaling, etc.) and the performance criteria (type of equilibrium achieved at the steady state, convergence speed, etc.).

As further work in this direction, it should be remarked that existing results regarding the analysis of equilibrium in wireless networks strongly depend on the topology of the network and the assumptions on the channel's models. A complete general framework for the analysis of equilibria and learning dynamics adapted to time-varying topology networks is still an open problem.

Chapter 3

System Model

This chapter provides the basic network model and the notations used throughout this dissertation to represent a tactical DSCN. This model is based on an abstraction of the real system presented in Section 1.1.1. However, by imposing some extra constraints, it is possible to describe different types of *ad hoc* networks. For instance, it can easily represent networks in which all nodes are interested in communicating with the same receiver (i.e., where the CH acts also as a receiver) and networks in which the CH manages several point-to-point communications inside the DSCN similarly to a cellular communication.

In our model of DSCNs, radio devices are arranged into groups, to which we refer as clusters, and each cluster is managed by a central controller or a CH. In a tactical network, each cluster may represent a national entity, or a particular subset of devices that are in-range and can communicate with each other. The cluster formation and the CH selection functions are responsible for creating in real time the clusters and their heads [122]. In general, clusters are allowed to merge and split, depending on the needs of the mission, and each device may at any time become a CH. However, in this thesis, we assume the cluster formation and CH selection functions to be completed. The main task of the CH is to choose the logical channel in which its cluster must operate and to determine the power levels to be used by all radio devices inside the cluster. Hence, this network model is decentralized, in the sense that there exist several CHs autonomously taking decisions, and centralized, in the sense that radio devices inside a cluster implement the decision adopted by their corresponding CH.

3.1 System Details

Consider a DSCN in which all devices coexist within the same spectrum subject to mutual interference. In this network, for the sake of simplicity we assume the presence of

only intra-cluster single-hop communication. The assumptions are not very limiting in fact, in real military networks [20, 123], inter-cluster communications are handled by specific borderline devices belonging to two different clusters at the same time. In other words, communications between devices belonging in different clusters happen through multiple intra-cluster single-hop passages. Here, devices are arranged into groups, referred to as clusters. Each cluster is controlled by a CH that harmonizes the intra-cluster communications by strategically choosing a channel (e.g., a frequency band) and a power level to be used by all the nodes in the corresponding cluster. Two instances of such networks are depicted in Figure 3.1 and Figure 3.2. In both figures, crosses represent the transmitters, circles represent the receivers and different color differentiate devices belonging to different clusters.

Figure 3.1 represents a static dense DSCN, in which a certain number of static clusters share a rather limited amount of logical channels. Figure 3.2 represents a mobile network in which a mobile cluster, the one on the bottom of the figure, moves at a constant speed towards the four static clusters.

For this networks, let $\mathcal{K} = \{1, 2, \ldots, K\}$ be a set of K clusters. Let also $\mathcal{L}_k = \{\ell_{1,k}, \ell_{2,k}, \ldots, \ell_{k,L_k}\}$ denote the set of L_k links within cluster k, with $k \in \mathcal{K}$. Each link is composed of a transmitter and a receiver. For the sake of simplicity, this role is assumed to be time-invariant. The set of all the links in the network is denoted by $\mathcal{L} = \cup_{k \in \mathcal{K}} \mathcal{L}_k$, with $L = |\mathcal{L}|$ the total number of links in the network.

Let $\mathcal{C} = \{1, 2, \ldots, C\}$ be the set of C channels into which the total spectrum is divided. All channel gains are assumed to be time-invariant for the whole duration of one transmission. Cluster k uses only one channel denoted by $c_k \in \mathcal{C}$ and a transmit power level p_k that is chosen from a finite set $\mathcal{P} = \{0, \ldots, P_{\max}\}$ of $Q = |\mathcal{P}|$ power levels. The maximum transmittable power level is denoted by $P_{\max}$ and it is assumed to be the same for all clusters. This model, or some minor variations, has been used in several works [63, 67, 68, 78, 107, 124–126], and high fidelity simulations have validated its results [20].

A pair of a channel and a power level is referred to as an *action*, i.e, $a_k = (c_k, p_k) \in \mathcal{A}$, where $\mathcal{A} = \mathcal{C} \times \mathcal{P}$ is the set of actions. The vector describing the whole network configuration is denoted by $\boldsymbol{a} = (a_1, a_2, \ldots, a_K) \in \mathcal{A} \times \ldots \times \mathcal{A} = \mathcal{A}^K$, and it is often referred to as an action profile.

The goal is to design a fully decentralized algorithm that selects a network configuration vector $\boldsymbol{a}^* \in \mathcal{A}^K$ that is a solution of the following optimization problem:

$$\begin{cases} \max\limits_{\boldsymbol{a} \in \mathcal{A}^K} \sum\limits_{k=1}^{K} \varphi_k(\boldsymbol{a}) \\ \text{s.t. } \xi_\ell(\boldsymbol{a}) > \Gamma \qquad \forall \ell \in \mathcal{L}^*. \end{cases} \tag{3.1}$$

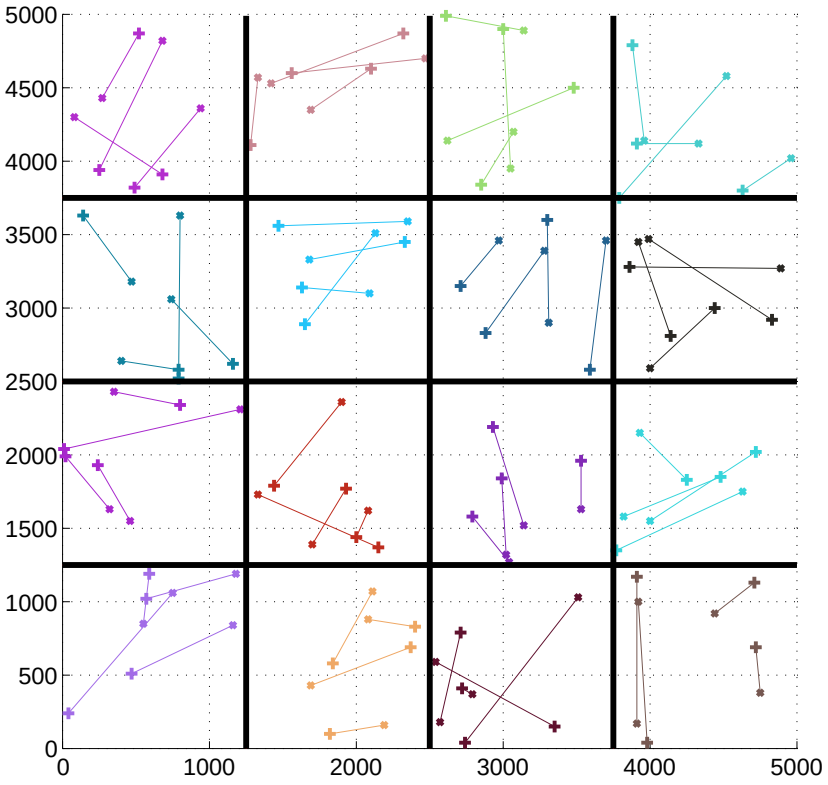

FIGURE 3.1: A 5 km $\times$ 5 km square field divided into $K = 16$ clusters. Devices are positioned randomly inside each cluster.

The function $\varphi_k : \mathcal{A}^K \to [0, 1]$ determines the performance $\varphi_k(\boldsymbol{a})$ achieved by the cluster k when the actions chosen by all clusters correspond to the action profile $\boldsymbol{a}$. The function $\xi_\ell(\cdot) : \mathcal{A}^K \to [0, 1]$ represents the QoS constraints to which link ℓ is subject, and Γ represents the minimum QoS a link must obtain. The set $\mathcal{L}^* \subseteq \mathcal{L}$ is defined as the largest set of links for which the constraints in (3.1) can be simultaneously satisfied. Note that $\mathcal{L}^*$ depends on all the individual constraints that are autonomously determined by each link. Thus, not all the constraints might be simultaneously satisfiable. Fixing the set $\mathcal{L}^*$ is a mathematical maneuver put in place in order to guarantee that the optimization domain in (3.1) is not empty. Later, it is shown that there is no loss of

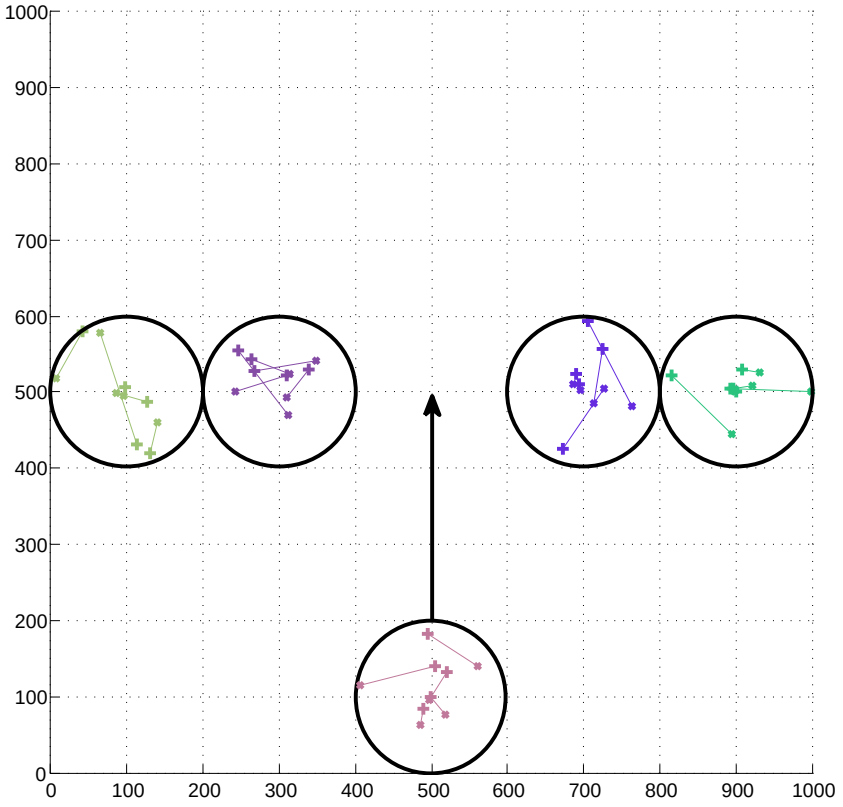

FIGURE 3.2: Cluster positions at the beginning of the mobility scenario with $K = 5$ clusters in a field of 1 km side. Four clusters are static and aligned, the cluster at the bottom is the one in mobility.

generality by assuming the set $\mathcal{L}^*$ to be known in advance. The formulation in (3.1) might describe a large set of network optimization problems that do not necessarily need to be convex. For instance, by properly selecting the functions φ_k and ξ_ℓ, it is possible to analyze problems such as: (a) the throughput maximization problem subject to particular delay constraints; (b) the transmit power minimization subject to a particular network reliability constraint; and other problems.

Here, the final goal is to design a decentralized behavioral rule that allows the network to achieve an operating point $\boldsymbol{a}^*$ that is a solution of (3.1) based only on local intra-cluster available information.

3.2 Particular Case

In this section, a case of particular interest of the system presented in Section 3.1 is presented. In the following of this thesis, this model will be used in order to validate the theoretical conclusions. In this system,the goal is to minimize the total power used by the network, while guaranteeing a certain level of QoS defined as a minimum SINR.

This translates in setting the following:

$$\begin{cases} \varphi_k(\boldsymbol{a}) = 1 - \frac{p_k}{P_{MAX}} \\ \xi_\ell(\boldsymbol{a}) = \mathrm{SINR}_\ell(\boldsymbol{a}). \end{cases} \tag{3.2}$$

In order to evaluate the SINR value, let us assume that the link ℓ belongs to the cluster k that selected the channel c for its transmission. Hence, the SINR level is expressed by:

$$\mathrm{SINR}_\ell = \frac{p_k g^{(c)}_{(\ell,\ell)}}{\sigma^2 + \mathrm{MAI}_{(\ell)}}, \tag{3.3}$$

where $\mathrm{MAI}_{(\ell)}$ represents the multiple access interference (MAI) suffered by the receiver of the link ℓ, $g^{(c)}_{(\ell,\ell)}$ indicates the channel power gain between the transmitter and the receiver of the ℓ-th link and as usual σ^2 denotes the thermal noise variance at the receiver. The MAI is evaluated as:

$$\mathrm{MAI}_\ell = \sum_{j \in \mathcal{K} \setminus k} \mathbb{1}_{\{c=c_j\}} \sum_{\ell \in \mathcal{L}_j} p_j g^{(c)}_{(m,\ell)}, \tag{3.4}$$

where $g^{(c)}_{(m,\ell))}$ denotes the channel power gain between the transmitting device of the link m, which is assumed belonging to the cluster $j \neq k$, and the receiving device of the link ℓ, while $\mathbb{1}_{\{\}}$ is the standard indicator function.

In this thesis, two different kinds of channel gains are considered: block fading channels and Rayleigh fading channels. In the first case, channels' power gain is both time and frequency invariant for the duration of one transmission and depends only on the distance between transmitters and receivers. In the second case, the path-loss power attenuation is at each time instant multiplied for the realization of a Chi-square random variable. Hence, the power attenuation between the transmitter of the m-th the link and the receiver of the ℓ-th link is given by [127]:

$$g^{(c)}_{(m,\ell)} = \rho^2_{(c)} \frac{G_m G_\ell h_m^2 h_\ell^2}{d^4_{(m,\ell)}}, \tag{3.5}$$

where, G_m and G_ℓ represent the antenna gains, h_m and h_ℓ the height of the antennas, $d_{(m,\ell)}$ is the distance between the two devices, and $\rho_{(c)}$ is the realization of a stochastic process distributed according to a Rayleigh distribution.

3.3 Game Model

The purpose of this section is to present a game theoretical model of the system presented in Section 3.1. The normal-form game of the system is represented by the triplet:

$$\mathcal{G} = \left(\mathcal{K}, \{\mathcal{A}_k\}_{k\in\mathcal{K}}, \{u_k\}_{k\in\mathcal{K}}\right). \tag{3.6}$$

The set $\mathcal{K}$ represents the players, i.e., the K CHs in the network; the set $\mathcal{A}$ represents the individual actions of all players. Note that all players have the same set of actions. An action of player k, denoted by $a_k = (c_k, p_k) \in \mathcal{A} = \mathcal{C} \times \mathcal{P}$, is a pair made of a logical channel index and the transmit power level to be used by all links inside the corresponding cluster. We design the utility function of player k, $u_k : \mathcal{A}_k \to [0, 1]$ as:

$$u_k(\boldsymbol{a}) = \frac{1}{1 + \beta L_{\max}} \left(\varphi_k(\boldsymbol{a}) + \beta \sum_{\ell\in\mathcal{L}_k} \mathbb{1}_{\{\xi_\ell(\boldsymbol{a})>\Gamma\}} \right), \tag{3.7}$$

where β is a design parameter that balances the tradeoff between the number of links that can be satisfied $\sum_{\ell\in\mathcal{L}_k} \mathbb{1}_{\{\xi_\ell(\boldsymbol{a})>\Gamma\}}$, and the maximization of the function φ_k. This utility function is designed in order to have some useful features explained in the following.

- The utility function (3.7) is monotonically increasing with the number of links that are able to satisfy their individual constraints inside the corresponding cluster k, and with the value of the function φ_k that determines the global performance of cluster k.
- As it will be shown in Theorem 4.2 in Section 4.1, for a particular choice of the parameter β, i.e., $\beta > K$, the stochastically stable points of the TE learning algorithm introduced in Section 2.2.4.1 are both NE equilibria of the game $\mathcal{G}$ and solutions of the optimization problem in (3.1).
- As long as each link can locally evaluate their own QoS measure $\xi_\ell(\cdot)$, each CH can compute the value of (3.7) with only intra-cluster available information, avoiding the need for inter-cluster information exchange.
- The value that needs to be fedback from the links to the CH can be transmitted with only one bit per link per algorithm iteration, drastically reducing the level of overhead necessary for the communication.

Generally, the utility of each player in the game $\mathcal{G}$ depends on the whole action profile $\boldsymbol{a}$. As a consequence, in the following we assume that game $\mathcal{G}$ is an IG as defined in Definition 2.2. This is a reasonable assumption, since, physically, this means that no link is isolated from all the others.

In the particular case described in Section 3.2 the functions $\varphi(\cdot)$ and $\xi(\cdot)$ are defined by (3.2), hence the utility function becomes:

$$u_k(\boldsymbol{a}) = \frac{1}{1+\beta L_{\max}} \left(1 - \frac{p_k}{P_{\text{MAX}}} + \beta \sum_{\ell \in \mathcal{L}_k} \mathbb{1}_{\{\text{SINR}\ell(\boldsymbol{a}) > \Gamma\}} \right). \tag{3.8}$$

By simple inspection, it can be noticed how the parameter β balances between the CHs interest to save power (lower values of β) and to increase the chances of satisfying the SINR constraints for the maximum possible amount of links in the cluster. Notice that in order to evaluate $\mathbb{1}_{\{\text{SINR}\ell(\boldsymbol{a})>\Gamma\}}$ the links can adopt both direct methods, such as an estimation of the SINR though pilots, and indirect method such an ACK/NACK system based on a cyclic redundancy check (CRC) as shown in [126].

3.4 Closing Remarks

This chapter presented a full mathematical characterization of a general DSCN. The DSCN has been modeled as a clustered *ad hoc* network in which each cluster is an autonomous entity managed by a CH. The CH fulfills the purpose of managing the intra cluster communications and to choose the transmitting channels and the power levell for all the transmitters inside the cluster. The abstraction of a DSCN is based on the following assumptions:

- Devices have a fixed role, transmitters or receivers;
- The number of devices and clusters does not change;
- Transmitters wish to transmit the whole time (high load);
- Communication happens only inside clusters;
- Communication is always single-hop.

The global performance for the network is expressed through an optimization problem. Therefore, a designer can chose the goal of the network by properly designing two functions: one describing the performance of the communications, and one expressing the constraints. A particular case in which the goal is to minimize the power used while maximizing the amount of successful transmissions is presented.

Furthermore, this chapter proposed a game model in normal-form of the abstraction above. For this game, a utility function showing particularly interesting features was designed. The main features of this utility function are that it can be evaluated from the CHs with only intra-cluster available information, and that, among the elements of

the NE set, those NE showing the highest social welfare coincides with the solution of the optimization problem.

Chapter 4

Applications and Results

This chapter presents the main results of this thesis on the algorithmic design for DSCNs. First, theoretical results regarding the converging points of the TE learning algorithm are provided. The link between the NE learned by this iterative process and the solution of the optimization problem described in Chaper 3 is assessed and discussed. The average number of iterations that algorithm needs to execute in order to reach an NE, and the probability of the algorithm to be at at NE are evaluated theoretically and validated through numerical simulations. Both the average number of iterations that the algorithm needs to reach an NE, and the probability of the algorithm to be at at NE depends on the experimentation frequency of the TE learning algorithm.

Further, this chapter analyzes and compares the performance of the iterative procedures introduced in Section 2.2. Different particular scenarios are used as testbed in order to assess the performance of each algorithm. The limitation in performance of the TE learning algorithm are individuated and a heuristic enhancement of the algorithm is designed and tested. The chapter close with a discussion on the performance of this algorithm with respect to various models of DSCNs.

4.1 Theoretical Results

This section presents the theoretical results pertaining to the convergence points and the speed of convergence of the TE learning algorithm introduced in Chapter 2. A strong connection between the solutions of the optimization problem in (3.1) and the NE of the game $\mathcal{G}$ introduced in Chapter 3 is established via the utility function (3.7).

4.1.1 Equilibrium Points

Theorem 4.1. *Let all the players of the game $\mathcal{G}$ implement the TE learning algorithm, and adopt the utility function in* (3.7)*. Let let $\beta \in \mathbb{R}$ satisfy $\beta > K$, and denote by $\mathcal{A}_{\text{NE}}$ the set of NE of the game $\mathcal{G}$, assumed non-empty. Denote by λ_n the number of links satisfied at the n-th NE, with $n \in \{1, \dots, |\mathcal{A}_{\text{NE}}|\}$ and let $\Lambda = \max_{n \in \{1,\dots,|\mathcal{A}_{\text{NE}}|\}} \lambda_n$. Then, the stochastically stable points of the TE learning algorithm are the NE in which there are at least Λ links that satisfy their individual constraints.*

The proof of this theorem is reported in Appendix B. Theorem 4.1 states that, if each player sets $\beta > K$, then the stochastically stable points of the TE learning algorithm are those NE with the largest possible number of links satisfying their constraints. Here, β represents the trade-off between the interest in satisfying the constraints for the largest set of links and the maximization of the sum of the objective functions. Intuitively, setting $\beta > K$ means that the designer has more interest in satisfying the QoS constarints than in maximizing the objective function. If one considers the system model in Section 3.2, this means that the designer has more interest in satisfying the SINR constraints even just for one link rather than saving the network power. However, at parity of link satisfied, the algorithm selects the solution in which the minimum power consumed.

The next theorem links this result with the global optimization problem in (3.1).

Theorem 4.2. *Let all the players of the game $\mathcal{G}$ implement the TE learning algorithm with the utility function in* (3.7)*, and let $\beta \in \mathbb{R}$ satisfy $\beta > K$. Let $\mathcal{A}^\dagger \subseteq \mathcal{A}^K$ be the set of solutions of the optimization problem in* (3.1)*, and let $\mathcal{L}^*$ be the largest set such that $\exists\ \boldsymbol{a} \in \mathcal{A}^\dagger$ and $\forall \ell \in \mathcal{L}^*,\ \xi_\ell(\boldsymbol{a}) > \Gamma$ and $|\mathcal{L}^*| = L^*$. Also let $\mathcal{A}_{\text{NE}}$ be the set of NE of the game $\mathcal{G}$, and assume $\mathcal{A}_{\text{NE}} \cap \mathcal{A}^\dagger$ is non-empty. Then, the TE algorithm is stochastically stable in an action profile $\boldsymbol{a}^*$ such that $\boldsymbol{a}^* \in \mathcal{A}_{\text{NE}} \cap \mathcal{A}^\dagger$.*

The proof of this theorem is reported in Appendix C. Note that the set of solutions of (3.1) is non-empty as long as there exists a set $\mathcal{L}^*$ such that the optimization domain is not an empty set. This theorem states that the stochastically stable points of the TE algorithm are those NE that maximize the sum of the network objective functions among the action profiles that satisfy the constraints for the largest possible set of links. For instance, if the network objective functions $\varphi_k(\cdot)$ are decreasing with respect to the power level p_k, then the stochastically stable points are those NE which satisfy the constraints for the largest number of links and minimize the power consumption. As a further example of the implications of this theorem, consider the system model in Section 3.2. In such a network, if β is set to a value greater than the number of clusters in the utility function (3.7), then the configuration set by the TE learning algorithm is with high probability the configuration where the largest possible set of links achieve simultaneously the target SINR Γ using the minimum power possible.

4.1.2 Convergence Time

This section studies the convergence properties of the TE algorithm in a particular scenario.

The TE learning algorithm defines a large discrete time Markov chain (DTMC) over the set of states. Studying the behavior of the algorithm on such a chain is a difficult problem due to the number of states, transitions and parameters. For this reason, a simplified version of the system model introduced in Section 3.2 is considered. This allows the estimation of the average number of time instants that are required to reach an NE for the first time and the expected fraction of time the system is at an NE action profile.

For the ease of the presentation, consider $L_k = 1$, i.e., each cell possesses only one link. Such a network is depicted in Figure 4.1. The functions φ and ξ are thus defined as:

$$\begin{cases} \varphi_k(\boldsymbol{a}) &= 1 - \frac{p_k}{P_{\text{MAX}}} \\ \xi_k(\boldsymbol{a}) &= \text{SINR}_k(\boldsymbol{a}). \end{cases} \tag{4.1}$$

In this particular formulation, the aim is to minimize the transmit power while keeping the SINR above a threshold Γ for the largest number of links. In (4.1), since there is only one link per cluster, the link index is the same as the cluster index. Therefore the SINR of link k is evaluated as:

$$\text{SINR}_k(\boldsymbol{a}) = \frac{p_k g_{k,k}^{(c_k)}}{\sigma^2 + \sum\limits_{\ell \in \mathcal{K} \setminus k} p_\ell g_{k,\ell}^{(c_\ell)} \mathbb{1}_{\{c_\ell = c_k\}}}, \tag{4.2}$$

where $g_{k,\ell}^{(c_k)}$ indicates the channel power gain between the transmitter of link k and the receiver of link ℓ over channel c_k; and σ^2 represents the noise power. This problem has also been studied in [78]. Note that it is possible for the receivers to evaluate the SINR through pilots and training sequences. In the following, it is assumed that the number of channels C is grater than the amount of clusters K, and that the channel gains follow the weak interference model as in [14]:

$$\begin{cases} g_{k,k}^{(c)} = 1 & \forall k \in \mathcal{K} \text{ and } \forall c \in \mathcal{C} \\ g_{j,k}^{(c)} = \frac{1}{2} & \forall k \in \mathcal{K} \text{ and } \forall j \in \mathcal{K} \setminus \{k\} \text{ and } \forall c \in \mathcal{C}. \end{cases} \tag{4.3}$$

In the light of the description in Section 2.2.4, if the number of players K is large enough the following can be stated:

- The fraction of time player k is either at *watchful* or *hopeful* state is negligible compared to the fraction of time it spends in *discontent* or *content* state;

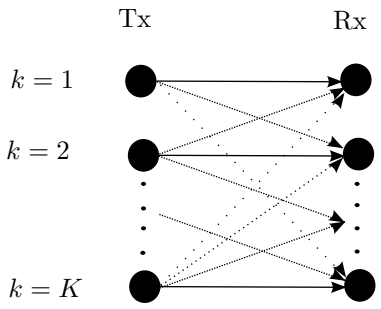

FIGURE 4.1: Simplified system model: symmetric parallel interference channel.

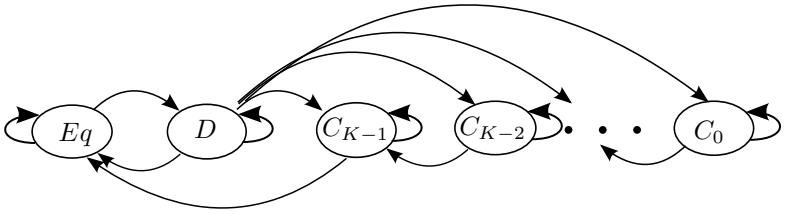

FIGURE 4.2: Markov chain describing the TE learning algorithm in the network. This model is used to study the convergence to the NE. The state Eq represents an NE action profile. C_{K-k} represents a state in which $K-k$ players are using an individually optimal action, D represents a state in which at least one player is *discontent*.

- At any time, the probability of having more than one player *discontent* is significantly lower than the probability of having only one or no *discontent* player.

In fact, in (2.11) the probability of accepting the outcome of the experimentation for a player which is discontent is close to one, moreover players do not adopt a *watchful* or *hopeful* state for more than one iteration. Section 4.1.3 shows that the theoretical results obtained on such a simplified model are good approximations also under less restrictive conditions as well.

Under these conditions, the resulting DTMC for studying the TE learning algorithm is represented in Figure 4.2.

In this figure, the final state represents an NE, the states labeled with C_{K-k} are those in which $K-k$ players use an individually optimal action and D a state in which one player is *discontent*. The transition probabilities are listed hereafter (the reasoning behind these transition probabilities is given in appendix D):

$$P(N,D) = \frac{K(K-1)^2\varepsilon^2}{C^2}\left(\frac{Q-1}{Q}\right)^2 \tag{4.4}$$

$$P(D,N) = \frac{(C-K+1)}{CQ} \tag{4.5}$$

$$P(D,C_{K-k}) = \frac{(C-K+k)}{C^k}\frac{(K-1)!}{(K-k)!} \tag{4.6}$$

$$P\left(C_{K-k},C_{K-k-1}\right) = (K-k)\frac{C-k}{CQ}\varepsilon. \tag{4.7}$$

Here, $P(N,D)$ is the transition probability between an NE and a state in which one player is *discontent*; $P(D,N)$ is transition probability between a state in which one

player is *discontent* and an NE; $P(D, C_{K-k})$ is the transition probability between a state in which one player is *discontent* and a state in which $K-k$ players are using an individually optimal action; and $P\left(C_{K-k}, C_{K-k-1}\right)$ is the transition probability between a state in which $K-k$ players are using an individually optimal action and a state in which $K-k-1$ are doing the same. The analysis of this DTMC leads to state the following theorems.

Theorem 4.3. *Let K, C, Q, and ε be the number of players, the number of channels, the number of power levels and the experimentation parameter respectively. Assume $C \geq K$. Let $L_k = 1$ and let the channel power gains be given by* (4.3)*. Then, if all players implement the TE learning algorithm, the expected number of iterations needed to reach the NE for the first time, $\bar{T}_{NE}$, is bounded as follows:*

$$\bar{T}_{NE} \leq \frac{CQ}{\varepsilon\left(C-K\right)}\left(1+\log\left(\frac{K\left(C-K+1\right)}{C+1}\right)\right) \tag{4.8}$$

$$\bar{T}_{NE} \geq \frac{CQ}{\varepsilon\left(C-K\right)}\left(\gamma+\log\left(\frac{K\left(C-K\right)}{C}\right)\right); \tag{4.9}$$

where, $\gamma \simeq 0.577$ is the Euler-Mascheroni constant.

Note that the time needed to visit an NE for the first time is directly proportional to the dimension of the action set (i.e., $|\mathcal{A}| = CQ$) and inversely proportional to the experimentation probability ε.

Theorem 4.4. *Let K, C, Q, and ε be the number of players, the number of channels, the number of power levels and the experimentation parameter, respectively. Assume $C \geq K$, $L_k = 1$, and let also the channel power gains follow* (4.3)*. Then, if all players follow the TE learning algorithm the expected fraction of time the system is at an NE is:*

$$(1-\delta) \approx \frac{1}{1+P(N,D)T_{BNE}}, \tag{4.10}$$

where

$$\begin{aligned} T_{BNE} &\approx \sum_{k=1}^{K} P(D, C_{K-k}) T_{CNE}(k) + \frac{P(D,N)}{\left(1-P(D,D)\right)^2}, \\ T_{CNE}(k) &\approx \frac{CQ}{\varepsilon\left(C-K\right)}\left(\gamma+\log\left(\frac{K\left(C-k+1\right)}{C+1}\right)\right), \\ P(D,D) &= 1-P(D,N)-\sum_{k=1}^{K} P(D, C_{K-k}). \end{aligned}$$

Note that the frequency of using an NE, i.e., $(1-\delta)$ depends on $\frac{1}{\varepsilon^2}$ as in (4.4). This means that larger value of ε implies that the network is at a NE for shorter average time. The approximation is given by the fact that T_{BNE} is replaced by its upper bound. Intuitively, this result can be motivated as follows. A NE is a state that is stable to

unilateral deviation. Hence, to leave it it is necessary that at least two players attempt at the same time to change their action. Since the probability of experimentation is ε, the probability of leaving the NE is approximatively ε^2. Therefore, the time spent on an NE is proportional to $\frac{1}{\varepsilon^2}$.

These theorems show that the stability of the TE algorithm and the time it needs to visit an NE for the first time are inversely influenced by the experimentation probability. Lower values of ε increase stability while higher values increase the speed of convergence. This bring a dilemma in choosing the right value of the experimentation probability that must come from correctly assessing the tradeoff between the need for a stable solution, and the need for quickly reaching the NE and for promptly responding to changes in the network.

Consider for instance the network in Section 3.2 with Rayleigh fading channels. The modification of the gains value imposes a modification of the NE. As a consequence lower values of the experimentation probability would make the algorithm too *conservative* forcing the clusters to use strongly suboptimal channels. On the other hand, if one considers block fading channels, too large values of the experimentation probability would make the algorithm change the network's configuration too fast, with consequent loss of performance.

4.1.3 Numerical Validation

Theorems 4.3 and 4.4 allow the calculation of the fraction of time the system uses an NE and the average number of iterations needed before visiting an NE for the first time, as a function of several design parameters, assuming the channel model expressed in (4.3). The following shows that these results also hold under a more general formulation.

All experiments presented here are run on the scenario represented in Figure 4.1, with two different sets of parameters. The first set is composed of: $K = 3$, $C = 4$, $\varepsilon = 0.02$ and $6 \leq Q \leq 10$; the second one is composed of $K = 4$, $C = 5$, $\varepsilon = 0.02$ and $6 \leq Q \leq 10$. In the first experiment, the fraction of time the network is an NE is estimated by running 10^7 iterations under two different channel models: the simple channels expressed in (4.3) and a channel power gain randomly drawn from a Rayleigh distribution. These results are summarized in Figure 4.3. The dashed line and the continuous line correspond to the theoretical results with the first and the second set of parameters respectively. In both cases, the numerical results are close to the theoretical lines showing the accuracy of the theoretical analysis. Notice that the theoretical line overestimates the fraction of time the network is at NE when the channels are subject to fading. The reason behind this is that since fading tends to change the NE the network may sometimes leave a NE state as a result of the fading.

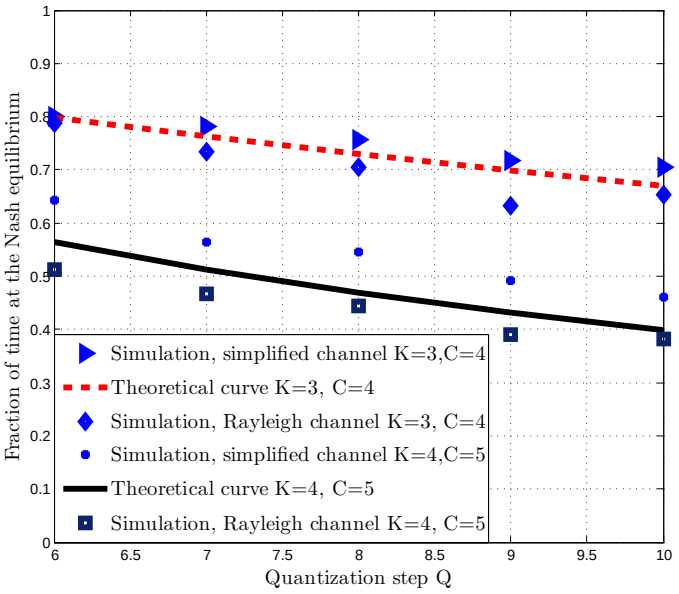

FIGURE 4.3: Fraction of time the system is at an NE, with the TE learning algorithm, $\varepsilon = 0.01$ and uniform probability distribution over the action set. Theoretical results are represented by the continuous lines, simulation results are represented by the markers for two sets of data and different channels: Rayleigh and the model in (4.3).

In the second experiment, the number of iterations needed to visit an NE for the first time is estimated and compared with the analytical results in Figure 4.4. This quantity is also an evaluation of how much the algorithm is responsive to change in the network. Predictably, increasing the dimension of the action set, i.e., increasing the amount of available channels C or quantization steps Q, brings slower convergence rates since the algorithm requires more time to explore all the possibilities. Notice that, while the lower bound appears to be a loose estimation of the numerical simulations results, the higher bound behaves as a good approximation of the actual values.

4.2 Asymptotic learning algorithms comparisons

In this section, we study and compare the asymptotic learning algorithms introduced in Section 2.2.2. The testbed is defined by the system model presented in Chapter 3, where the number of clusters is limited to $K = 2$ and in each cluster only one link is present, i.e., $\mathcal{L}_k = 1$.

Figure 4.5 reports the average spectral efficiency of the network as a function of the SINR , in the case where only 2 orthogonal channels are available. Here, all the algorithms iterate the same number of times (40 iterations). The difference in performance does not depend on the number of iteration as witnessed from Figure 4.7 that reports the

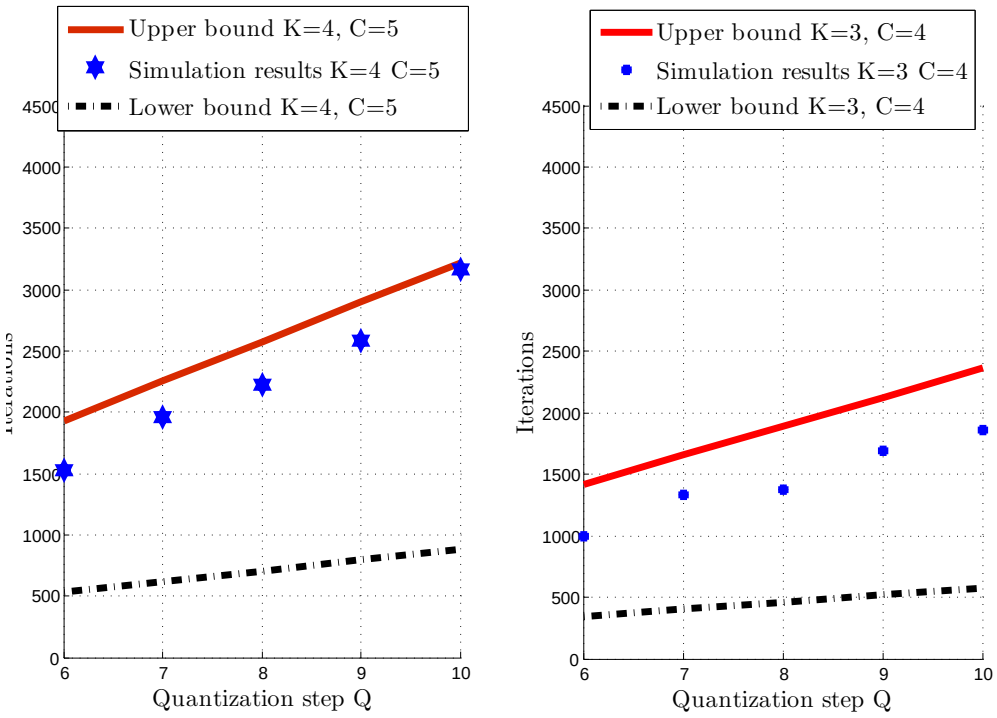

FIGURE 4.4: Number of iterations needed for the TE learning algorithm to visit an NE for the first time, with $\varepsilon = 0.01$ and uniform probability distribution on the actions set. The continuous lines represent (4.8), the dashed lines represent (4.9).

network spectral efficiency of the algorithms as a function of the number of iterations. From this figure one can see that the algorithm's performance remain mostly unvaried after 40 iterations. Predictably RM is the most performing learning algorithm directly followed by the FP and SFP. Concluding however that the RM is the best algorithm would be imprudent, since, in order to function, the RM learning algorithm to require more information on the game than any other algorithm considered in this thesis. Other two considerations are in order. First, the difference in performance between RM and FP or SFP is not due to the performance of the steady state rather on the speed of convergence. Second, even though it demands the same information as SFP and FP, in this settings the BRD is worst performing algorithms. The bad performance of the (simultaneous) BRD can be explained with the lack of convergence. Whenever the transmitters begin the learning procedure on the same channel they enter in an infinite loop in which they always collide. These two conclusions can be also inferred from Figure 4.6 where the trajectories of the algorithms are depicted.

This figures reports for each player the probability of choosing the first channel. Each blue point represents one iteration of the algorithms, the green point the converging state and the crosses represent the NE. In the plot on the top left, the trajectories of the BRD are reported. The two transmitters repeatedly select synchronously the same

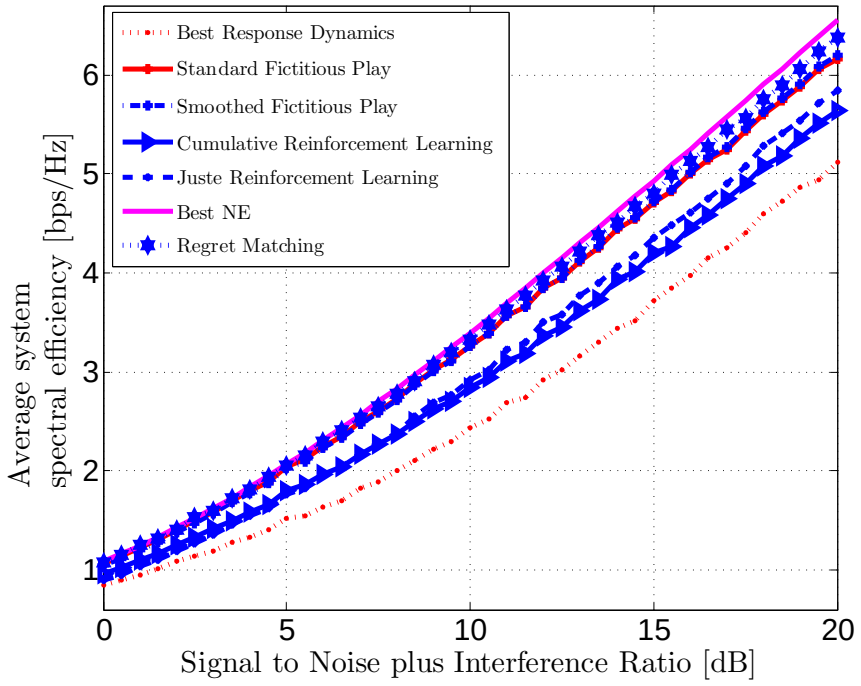

FIGURE 4.5: Average system spectral efficiency [bps/Hz] as a function of SINR with 40 iterations for the 2 players and 2 channel case.

channel entering in a cycle where they never exploit the free channel. The difference of time needed to converge to NE between FP and RM can be estimated by the amount of blue points present in the figure. In this figure it is also possible to see how FP and SFP converge to the best performing NE [1] while RL converges fast to a steady point that has no game theoretical meaning. In the trajectory of JUSTE-RL , it is possible to notice that, for this particular channel realization, it converges to the best performing NE,

To show the variation of the performance of the algorithm with respect to availability of resources, a similar experiment in which the channel are increased to to $C = 4$ is run. Figure 4.8 reports the result of this simulation. Interestingly, FP, SFP and RM always converge very close to the best NE. Nonetheless, this performance is achieved at the cost of a lot of information about the game. In particular, note that RL and JUSTE-RL are less performing, but at the same time, less demanding in terms of information. In this case the performance of the BRD is superior tot he one of the RL

[1] More precisely, SFP converges to an ε approximation of the NE. However ε is set small enough to make little difference.

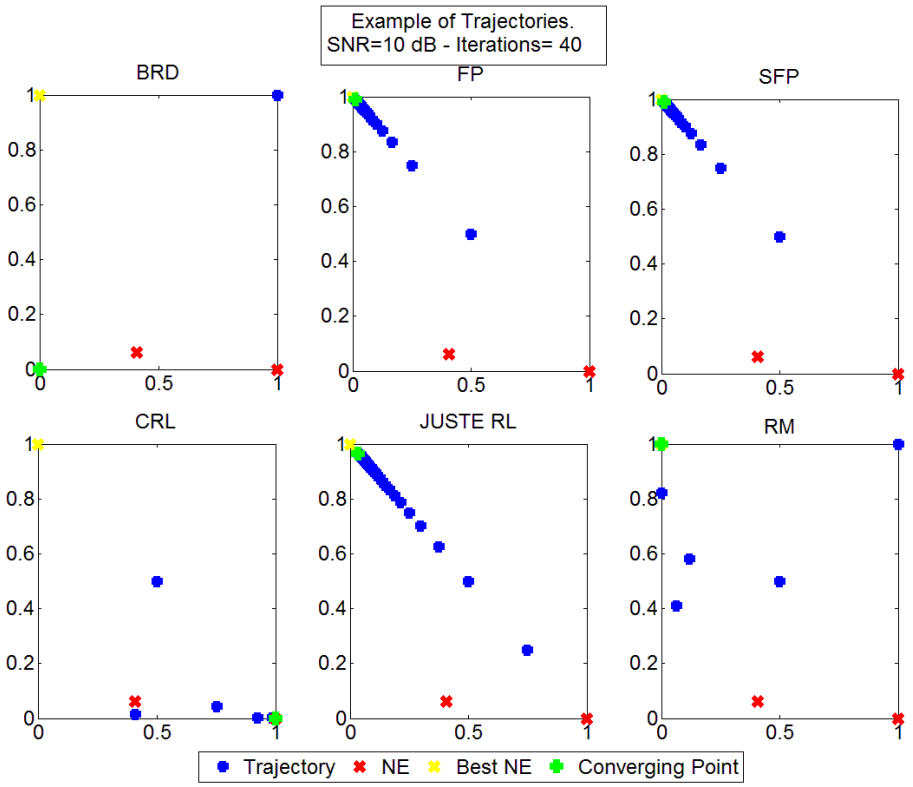

FIGURE 4.6: Example of trajectories. BRD bounces between unstable solution; FP and SFP converge close to the best NE; RL converges to a low performing NE, JUSTE-RL converges close to the best NE, RM converges close to the best NE.

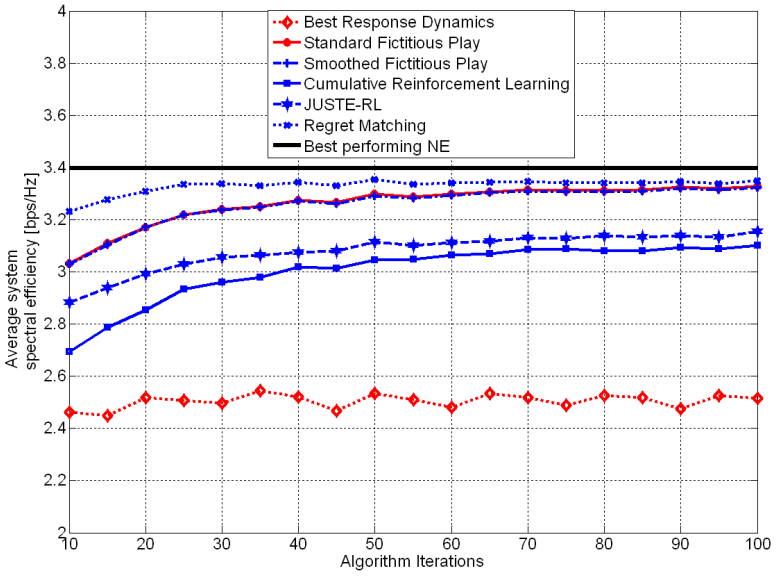

FIGURE 4.7: Average system spectral efficiency [bps/Hz] as a function of the number of iterations at a fixed SINR of 10 dB for the 2 players and 2 channel case.

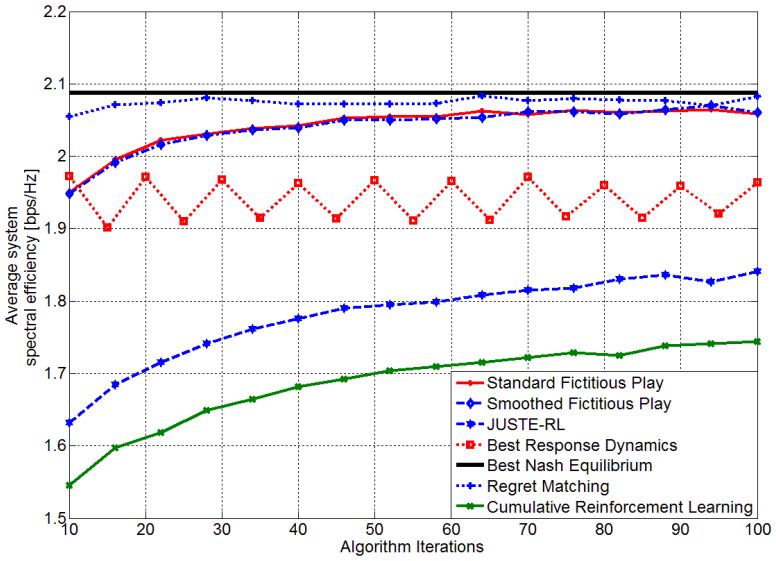

FIGURE 4.8: Average system spectral efficiency [bps/Hz] as a function of the number of iterations at a fixed SINR of 10 dB for the 2 players and 4 channel case.

and JUSTE-RL. This improvement is due to the fact that in this case the availability of resource allow for the algorithm convergence.

In order to evaluate how the availability of channels influence the performance o the different algorithms, we test them varying the number of channels. Figure 4.9 reports the results of this simulation. Here, the negative slope of the curves is due to the fact that we increase the number of available channels but transmitters remain subject to use a single channel. Hence, being $C > K$, there always exists a number of unused channels. The main observation in this figure is the following, the BRD becomes a very efficient solution when the number of channels is high enough to make the bouncing effect a very unlikely event. Conversely, JUSTE-RL exhibits a lower performance when the number of possible actions increases. This is basically because, in JUSTE-RL, each player plays all its actions with non-zero probability, in order to improve its utility estimation. This immediately implies that a growing set of actions increases the time spent trying suboptimal actions.

4.3 Trial and error performance

In this section, we evaluate the performance of the TE learning algorithm in configuring a DSCN. The metrics used for the evaluation are following:

- Average satisfaction (AS): The average number of times a link satisfies its SINR constraints.

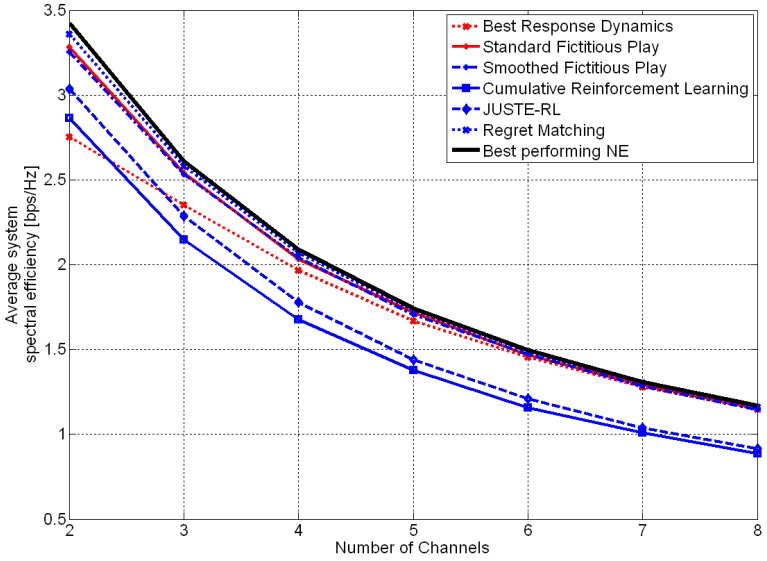

FIGURE 4.9: Average system spectral efficiency as a function of the number of channels, with SINR=10dB and 40 iterations.

- Average power consumption (APC): It is defined as the average amount of power used by the transmitters in a cluster to achieve the corresponding satisfaction level. It captures how much power is consumed per cluster.

4.3.1 Static DSCN

In this section, we analyze the performance of the TE learning algorithm in a static dense scenario as the one depicted in Figure 3.1. A square field of 5 km per side populated with $K = 16$ equally dimensioned square clusters is considered. Each cluster has a side of $\frac{5}{4}$ km and contains $\mathcal{L}_k = 4$ randomly positioned links. Each CH selects one over $C = 5$ available channels, and the minimum SINR level assumed for a receiver is an $\Gamma = 10$ dB.

In Figure 4.10, the AS in the network and the APC are plotted as functions of the TE iterations. The scarcity of resources in the network (i.e., the number of channels for cluster available) does not allow for full satisfaction, as consequence only an AS of 0.7 is achieved. Intuitively, this happens because in a network with $K = 16$ clusters sharing $C = 5$ channels, each cluster has on average two neighbor clusters that use its same channel.

In order to evaluate the optimum number of channels, TE's performance as a function of the available channels are evaluated. Available channels quantity varies between 4 and 18, and, for each of these values, 20 tests composed of 6000 TE iterations are run. The result is depicted in Figure 4.11. By simple inspection, one can notice

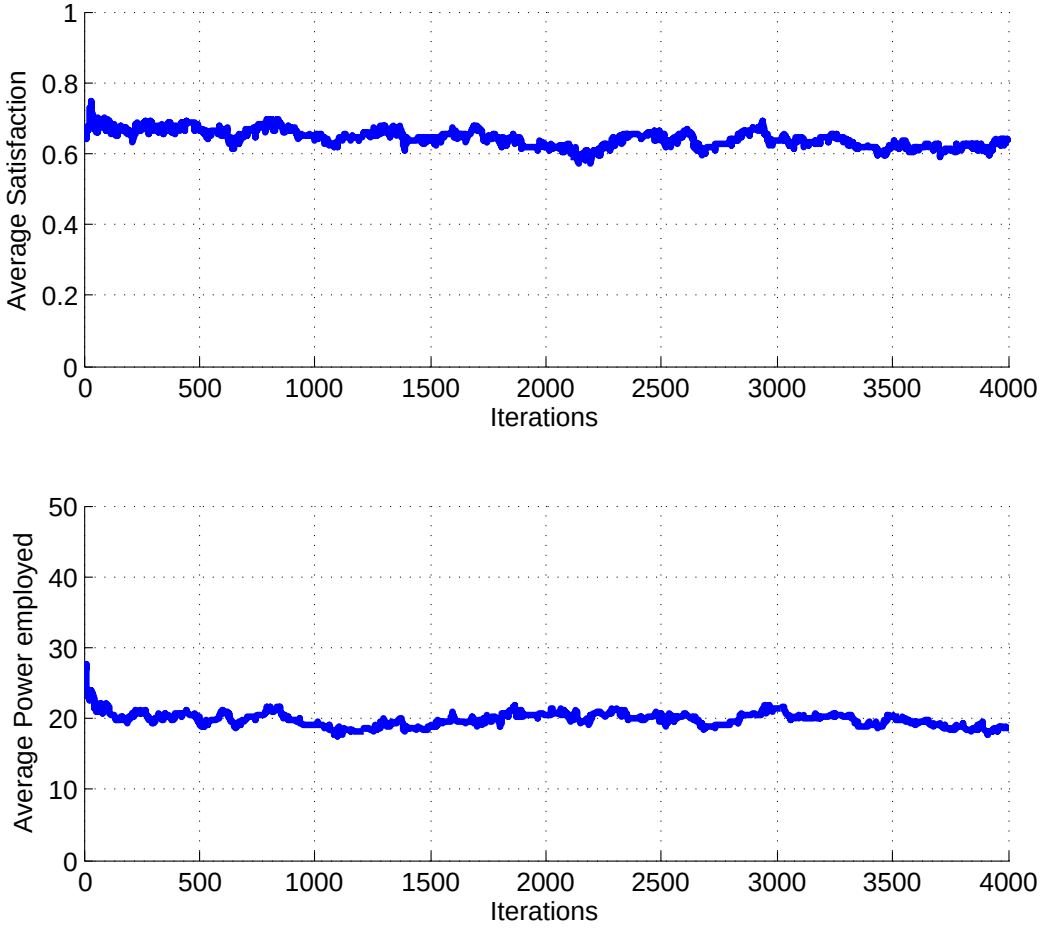

FIGURE 4.10: Achieved AS and APC as a function of the TE iterations for a square static scenario, with SINR-based feedback.

that the stochastic nature of the algorithm does not allow for full satisfaction even in presence of enough resources.

4.3.2 Mobile DSCN

This section evaluates the performance of TE in DSCNs in a mobile network, i.e., a network in which clusters are allowed to move. Assume $K = 4$ clusters to be aligned and sharing the spectrum while a fifth cluster is far away enough to be creating little interference. An instance of this starting situation is depicted in Figure 3.2. The fifth cluster begins to move at a constant speed towards the top of the field after 1500 iterations, and reaches the other four clusters after 2250 iterations, to reach the end of the field after 3000 iterations. Each CH selects one over $C = 2$ available channels, and the minimum QoS level assumed for a receiver is an SINR $\Gamma = 10$ dB.

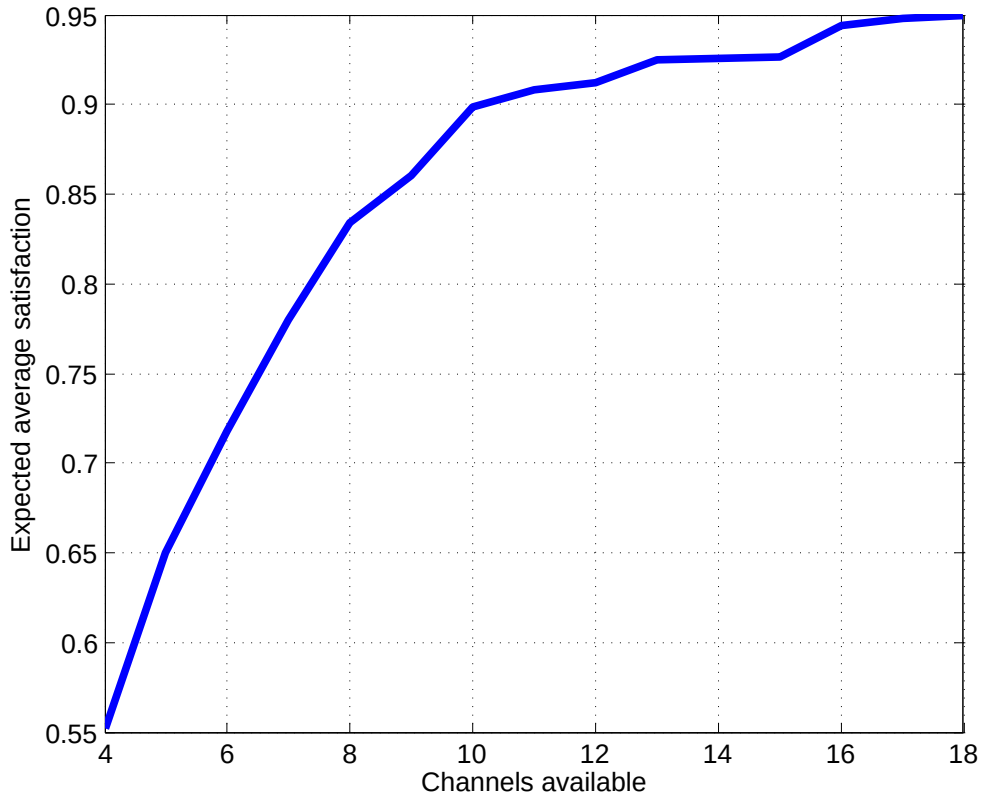

FIGURE 4.11: Expected satisfaction versus available channels. This plot has been realized assuming a square field as the one described in 3.2.

In Figure 4.12, we plot the global performance of the system in terms of AS and APC. The drop of system performance after 2000 iterations is due to the vicinity of the fifth cluster that increases the interference level. The algorithm reacts by increasing the power level and by modifying the channel configuration. The AS level, then, increases when the algorithm rearranges the channel and power allocation scheme in order to suit the new topology. Note that, when the mutual interference is too high, TE turns off one cluster by selecting zero power. The rationale behind this is that, if the desired level of SINR is not reachable by the current topological configuration, then the algorithm prefers to stop one of the clusters to improve the individual utility. When the algorithm reaches a different channel assignation pattern it is, again, possible to achieve a higher level of satisfaction.

Figure 4.13 reports a summary of the simulation run. Here each color represents one of the possible two channels, while the height of the bins represents the used power. The static clusters are indexed with numbers 1, 2, 4, and 5 while the moving cluster is indexed with the number 3. When the system reaches time instant (i) the 3rd cluster is close enough to create interference to the other clusters. This forces the system to reorganize the power-channel pattern. Comparing this figure with Figure 4.12, one can notice how the increased interference level provokes a drop in the AS. The algorithm reacts by increasing the power levels of the clusters and until the channels assignment

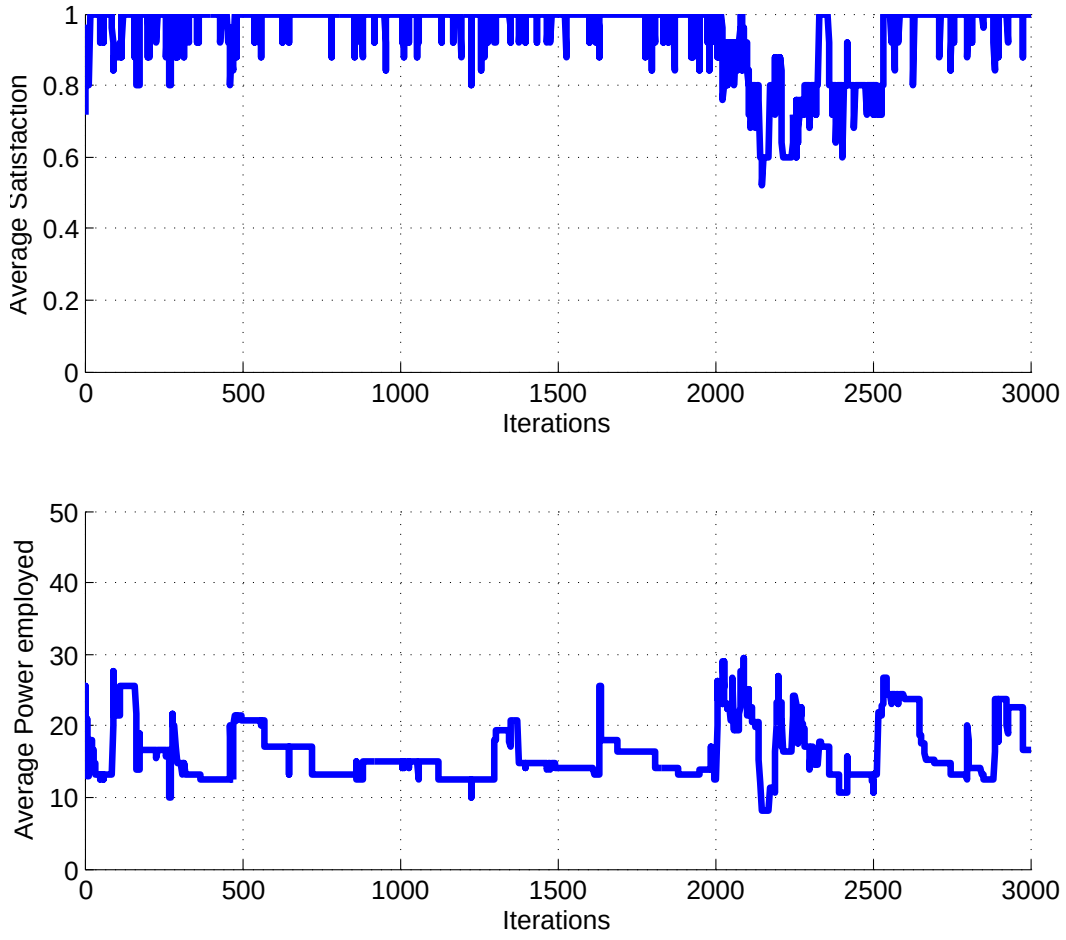

FIGURE 4.12: Achieved AS and APC as a function of the TE iterations for a the mobility scenario using the standard TE learning algorithm.

allow for lower power levels. That is, when the moving cluster is completely aligned with the others (*ii*) the system starts working in an orthogonal way and the power starts decreasing. At (*iii*) the cluster is far enough to stop creating interference.

On the down side, one can notice that the elevated experimentation factor forces the network to change the orthogonal configuration achieved around iteration 500.

4.3.3 Discussion

As highlighted by the previous results, even though TE has been shown to be capable of configuring a DSCN, its performance remain spoiled by an excess of instability due to the stochastic nature of the algorithm. In the next section we propose a heuristic solution to this problem, and we show the gain in performance.

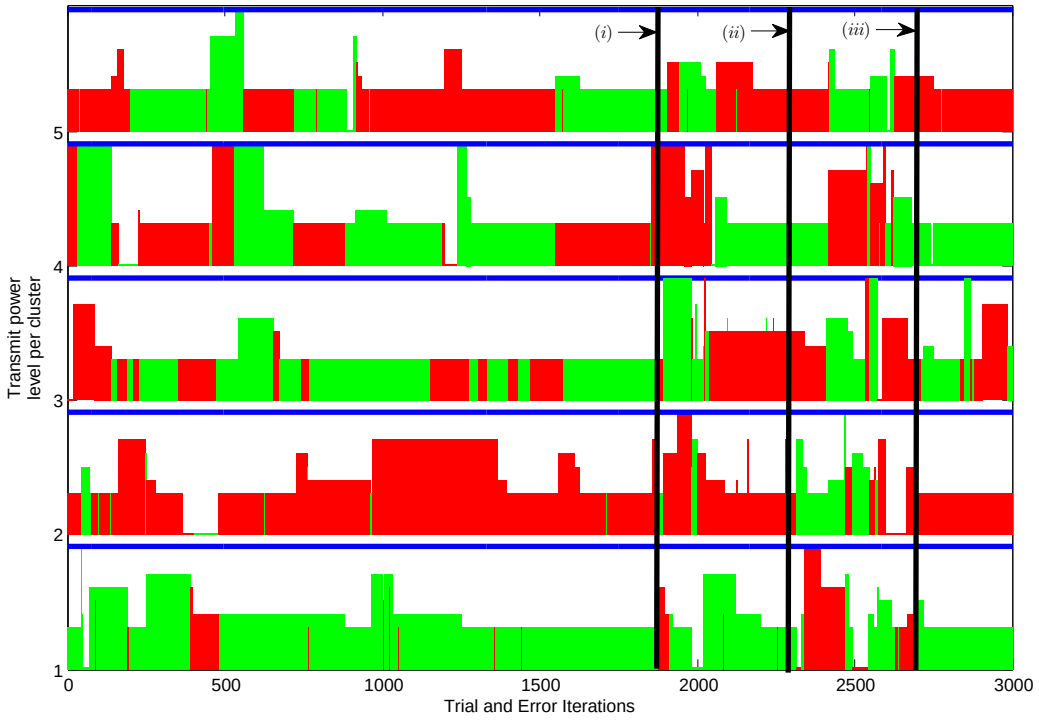

FIGURE 4.13: Channel-power allocation as a function of the TE iterations for the mobility scenario with two channels. Each color represents a different channel, and the heights of the graph the transmit power level. Clusters $1, 2, 4, 5$ are static, cluster 3 is in mobility. (i) beginning of the interference from the 3rd cluster, (ii) Five clusters are aligned, (iii) end of interference from the 3rd cluster. The blue solid lines represent $P_{\text{MAX}} = 50\text{W}$.

4.4 Optimal Dynamic Learning Performance

In this section, we compare the performance of the TE learning algorithm (a NE reaching algorithm) and ODL. Both algorithms share a state machine structure, a stochastic nature of the solution and they require the same amount of information. The main difference lies in the converging aspect. Implementing a social welfare maximizer may come at the cost of stability and of converging time. This can be considered as an instance of the exploitation versus exploration trade-off. That is, while the action profile selected with high probability by ODL has a higher social welfare than the NE selected by the TE learning algorithm, the time spent in learning in ODL is larger than the one spent in the TE learning algorithm.

In the next simulation, we run extensive experiments over the network described in section 3.2. A static network composed by $k = 16$ clusters is considered. The clusters share a spectrum composed by a variable number of logical channels, from $C = 2$ to $C = 7$. The results are represented in Figure 4.14. The comparison is performed in terms of social welfare overall the simulation time. The red dashed line represents the

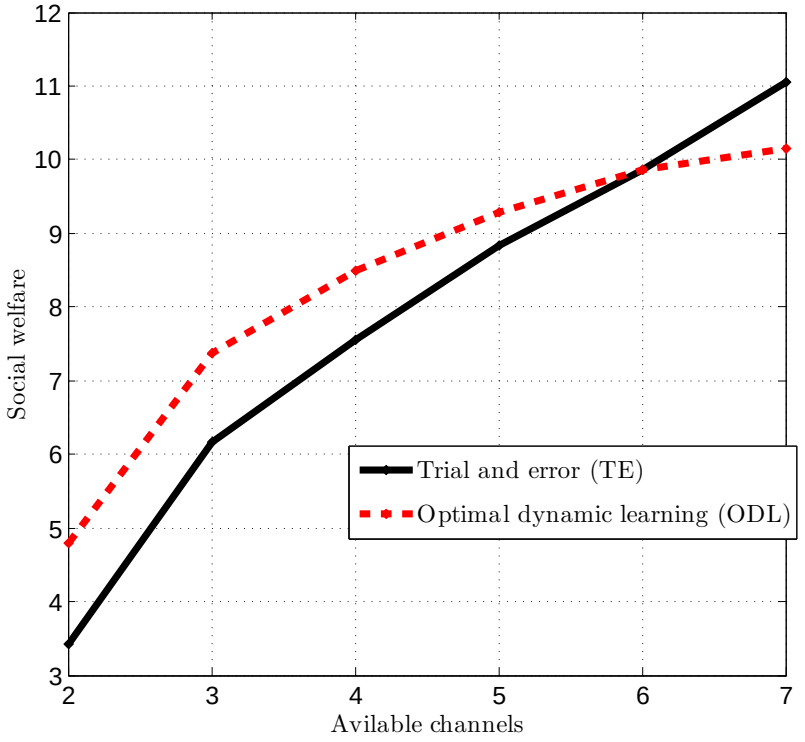

FIGURE 4.14: Comparison between NE-searching algorithm and ODL. The dashed curve represents the average social welfare obtained by ODL, the continuous curve represents the average social welfare obtained by TE both as a function of the available channels.

social welfare reached in the network when employing ODL, while the black continuous line represents the social welfare reached by employing TE. This plot shows that for such a network, ODL improves the performance only if $C \leq 6$. The reason behind is that, when the resources are scarce, the difference in performance between a Pareto optimal working point and a NE increases. As a consequence, under these conditions, the loss due to the instability of ODL is counterbalanced by the gain due to the selection of a well-performing working point. However, from a more practical standpoint, the gain in global performance brought by ODL is not sufficient to justify the use of ODL in real DSCN since stability of the solution is a primary target of any decentralized algorithm.

4.5 Enhanced Trial and error

Section 4.3 has shown that the main drawbacks of the TE learning algorithm are due to the instability of the action profile selected. Therefore, the purpose of this section is to presents enhanced trial and error (ETE), an enhanced version of the basic TE learning algorithm. The ETE learning algorithm's structure is first detailed in Section 4.5.1. The effect of the enhancement on the convergence capability is assessed in Section 4.5.2. Therefore, the algorithm is compared under various settings on the scenario described

in Section 3.2 against the original TE learning algorithm, in order to assess the improvements in terms of stability and performance in Section 4.5.3. Finally in Section 4.5.4, the algorithm's capabilities of configuring a DSCN are tested against other algorithms in the literature.

4.5.1 Enhanced Distribution and Settings

In its standard formulation, the TE learning algorithm [111] is characterized by a single time invariant ϵ and a uniform distribution over the whole action set. Motivated by the fact that experimentations on the set of channels brings higher instability than experimentations on the set of power levels, the experimentation is divided into two different steps. In detail, at each instant t, each player k in a *content* mood experiments with probability $\epsilon_c^k(t)$ a different channel and with probability $\epsilon_p^k(t)$ a different power level. This differentiation of the experimentations allows for an algorithm that experiments fast on the power levels in order to fast adapt to changes without wasting power and at the same time conserves a good channel-clusters association with low values of $\epsilon_c^k(t)$.

A second enhancement is given by turning the static ε of the original TE learning algorithm into time-varying values. This enhancement has the purpose of improving the flexibility of the algorithm that is allowed to modify the experimentation probability in accordance to the network's condition. That is, a network where the configuration is well performing will demand lower frequency of experimentations, while a fast changing network will demand higher experimentation frequency. CHs estimate the status of the network (static or fast varying) based on the amount of positive feedback they receive.

The evolution of $\epsilon_c^k(t)$ is given by the following rule:

$$\begin{cases} \epsilon_c^k(t) &= \max\left(\frac{\epsilon_c^k(t-1)}{2}, \epsilon_c^{min}\right) & \text{if } \sum_{\ell\in\mathcal{L}_k} \mathbb{1}_{\{\xi_\ell(\boldsymbol{a})>\Gamma\}} = |\mathcal{L}_k| \\ \epsilon_c^k(t) &= \epsilon_c^k(0) & \text{otherwise.} \end{cases} \tag{4.11}$$

In (4.11), $\epsilon_c^{min} > 0$ represents the minimum experimentation probability over the available channels and $\epsilon_c^k(0) > \epsilon_c^{min}$ represents the initial, maximum value. These parameters depend on the particular configuration of the system. Through numerical simulations, it has been found that some well-performing values are: $\epsilon_c^{min} = \frac{0.01}{K}$ and $\epsilon_c^k(0) = 0.01\frac{C}{K}$. Since no prior information is available on the channel gains, the experimentation on the channels follows a uniform distribution.

Each player k experiments a different power level with a constant probability ϵ_p^k. Such a probability is a uniform distribution over all the levels greater than p_k if $\sum_{n\in\mathcal{L}_k} \mathbb{1}_{\{\phi_n(\boldsymbol{a})>\Gamma\}} < |\mathcal{L}_k|$, whereas it is uniformly distributed over all the levels smaller than p_k, otherwise. Through extensive simulations, it has been found that a well-performing value is $\epsilon_p^k = 0.01\frac{C}{K}$.

When a player k is *discontent*, it experiments according to the following distribution:

$$\begin{cases} p_k(t) = P_{\max} & \text{with probability } \min\left(\frac{C}{K}, 1\right) \\ p_k(t) = 0 & \text{with probability } \max\left(1 - \frac{C}{K}, 0\right). \end{cases} \tag{4.12}$$

The rationale behind this is that any *discontent* player needs to test the network looking for a free channel. Clearly, the probability of finding a free channel increases with $\frac{C}{K}$. On the other hand, in the case in which no channel is free for transmission, zero power should be used to avoid wasting energy and creating interference.

4.5.2 Convergence to Nash Equilibrium

The following shows the effect of the enhancement on the stability and in the speed of the algorithm in reaching any stochastically stable point. A total of 10^4 iterations of TE learning algorithm are run with an underlying network as the one depicted in Figure 3.1, with $K = 4$ clusters each populated with one link, $C = 4$ channels, $Q = 5$ power levels and a target SINR of $\Gamma = 10$ dB. In Figure 4.15 the probability with which the TE learning algorithm selects an NE as a network action profile is plotted as a function of the experimentation probabilities ϵ_p and ϵ_c^{min}. Reducing the minimum experimentation probability on the channel sensibly decreases the instability of the system and thus increases the probability of the system of being at the NE. On the other hand, the stabilizing effect of reducing the experimentation probability on the power levels is balanced by the longer time that is needed for the system to reach an NE, as showed in Figure4.16. In this figure, the number of iterations used by the TE learning algorithm to reach, for the first time, an NE is plotted as a function of the experimentation probabilities ϵ_p and $\epsilon_c^{\min}$. Note that, the number of iterations needed to reach for the first time an NE represents also a measure of the speed of the algorithm to reach again an NE, once it is left. From a real-system implementation point of view, it is also an estimation of the ability of the algorithm to react to network changes that modify the NE set, e.g., fading, shadowing, mobility. By inspecting both plots, it appears that the experimentation frequency on the power levels should be relatively high, while the one on the channels should be relatively low with the exact optimal values depending on the other parameters of the network.

4.5.3 Comparison with Trial and Error Learning

In this section, the performance of the ETE learning algorithm is compared with the TE learning algorithm. The testbed scenario are composed of a static dense network and a mobile network.

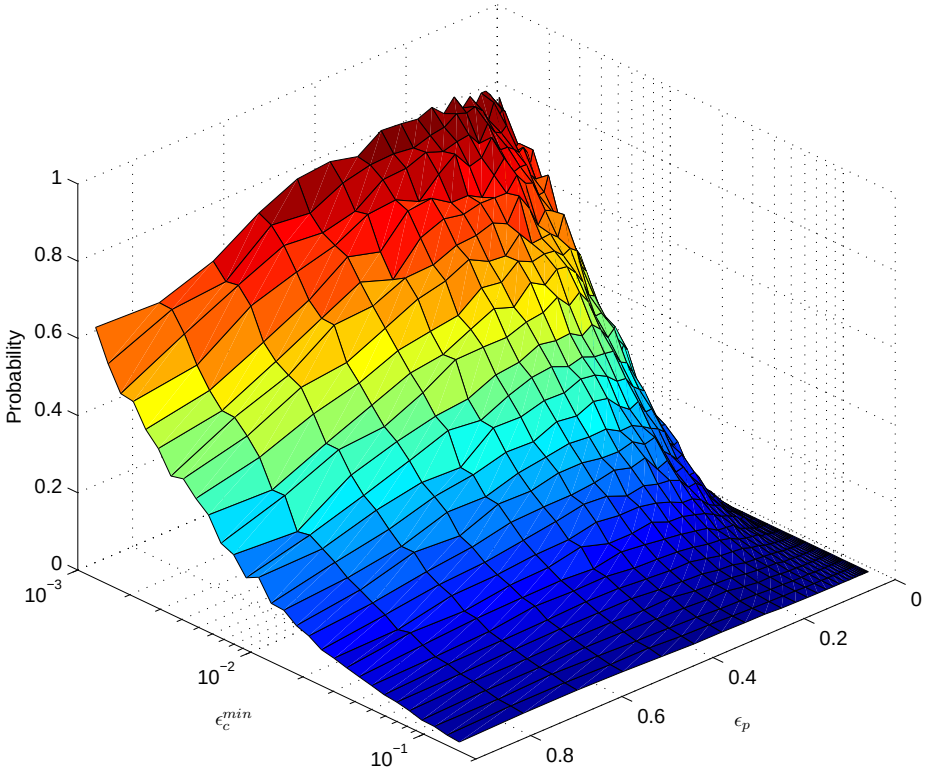

FIGURE 4.15: The plot represents the probability of observing the TE learning algorithm selecting an action profile which is an NE as a function of ϵ_p and $\epsilon_c^{\min}$. The underlying network is composed of $K = 4$ clusters, $L_k = 1$ links per cluster, $C = 4$ channels and $Q = 5$ power levels. The $\epsilon_c^{\min}$ values are reported in logarithmic scale.

4.5.3.1 Static DSCN

This section compares the performance of the ETE learning algorithms with the one of the TE learning algorithm in terms of AS and APC on the static dense network depicted in Figure 3.1. Both the case of block fading channels and the case of Rayleigh fading channels are considered. First, consider a static network composed of $K = 16$ clusters each with $N_k = 4$ links, $C = 5$ block fading channels, and the maximum power $P_{\max} = 50$W is quantized in $Q = 8$ levels. The results, reported in Figure 4.17, show that, in this case, for a similar amount of power spent, the ETE learning algorithm is able to satisfy more links. In particular, While TE achieves an AS of around 0.4 ETE achieves an AS of 0.6, satisfying almost the 20% more links for the same power consumed.

In the second experiment, the algorithms' performance are tested in presence of Rayleigh fading. The variance of the Rayleigh random process is set equal to 1. Also in this case, the improvement in terms of AS due to the enhancement is remarkable, though

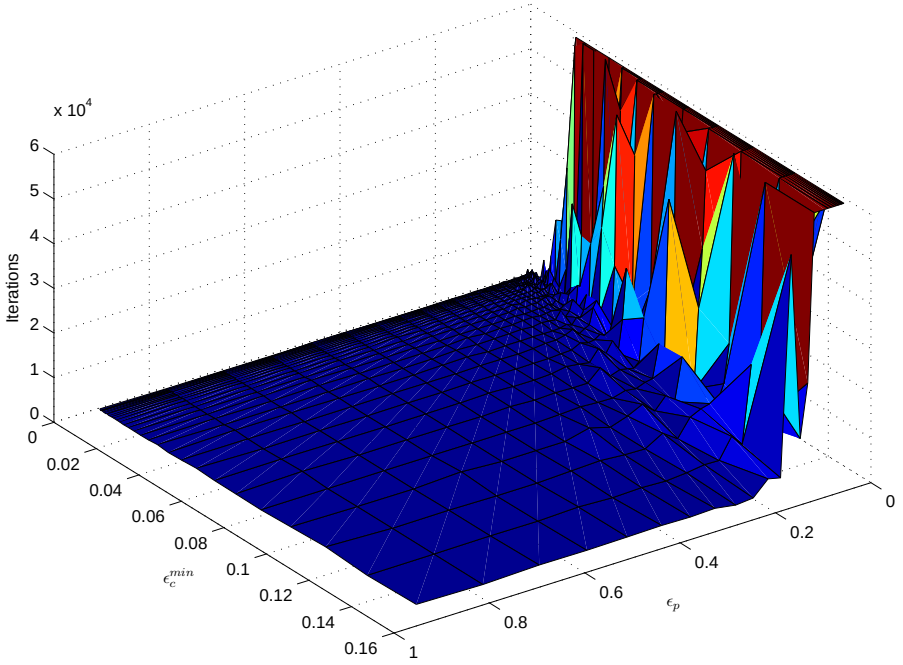

FIGURE 4.16: The plot represents the number of iterations between $t = 0$ and the instant in which an NE is played for the first time as a function of ϵ_p and $\epsilon_c^{\min}$. The underlying network is the same as in Figure 4.15.

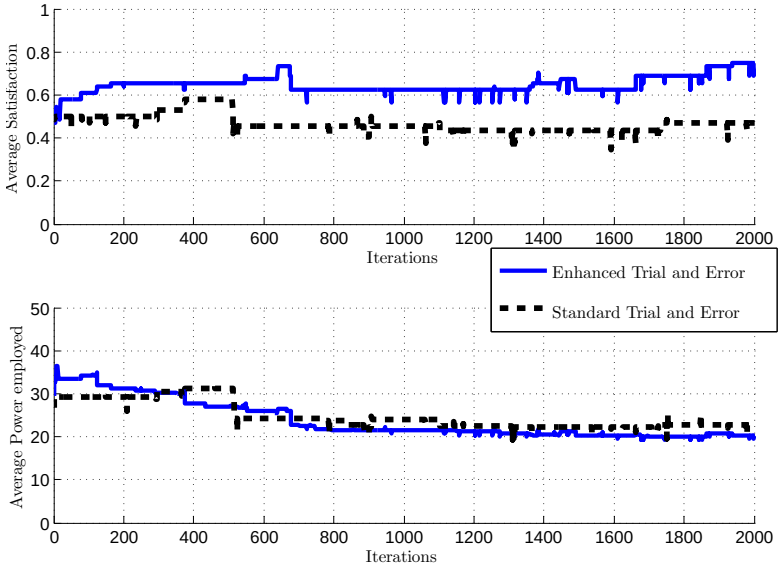

FIGURE 4.17: AS and APC in a static scenario in presence of block fading channels. The simulation parameters are the following: $\Gamma = 10$ dB, $C = 5$, 2000 iterations. The blue continuous line represents the ETE's AS (upper plot) and APC (lower plot), while the black dashed line the TE's AS (upper plot) and APC (lower plot).

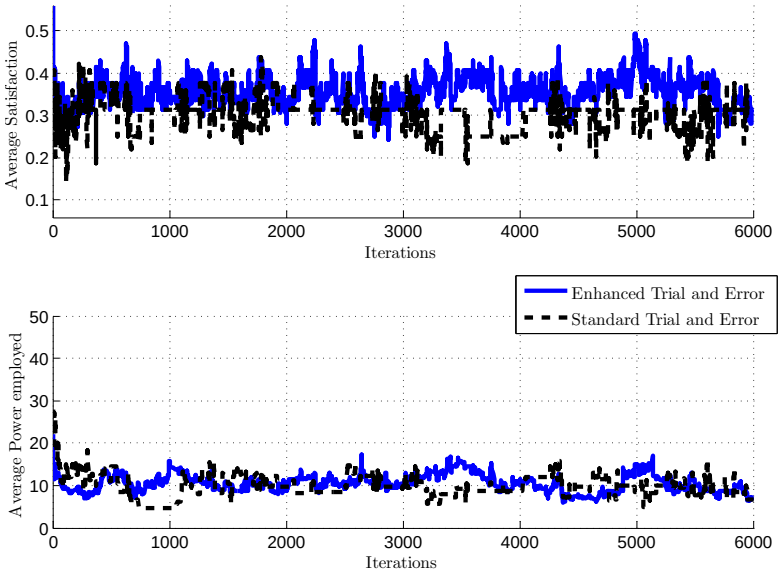

FIGURE 4.18: AS and APC in a static scenario with Rayleigh fading. The simulation parameters are the following: $\Gamma = 10$ dB, $C = 5$, 6000 iterations. The blue continuous line represents the ETE's AS (upper plot) and APC (lower plot), while the black dashed line the TE's AS (upper plot) and APC (lower plot).

less constant. In presence of fading, the working points of both algorithms are less stable. Indeed, the fast modifications in the channel gains imply equally fast variations on the optimal working points of the network. The conclusions of Theorem 2.10 holds also for the ETE learning algorithm, in fact both the TE and ETE learning algorithms hare the same state machine structure. As a consequence, both TE and ETE stochastically implement a globally optimal NE, hence, variations of these points modify the decision taken by both algorithms.

In the third experiment, 10000 iterations of both algorithms are run for different amounts (from $C = 2$ to $C = 18$) of available channels. For each simulation, Figure 4.19 reports the AS reached in the network, and the corresponding APC. The figure shows that the enhancement approximately provides the network a gain of one free channel, consuming a slightly lower level of power.

The next experiment aims at illustrating in detail the effect of the enhancement on the stability of the channel-cluster association. In order to do this 10000 simulations of both ETE and TE are run for different amount of available channels, from $C = 4$ to $C = 18$. The results are reported in Figure 4.20.

In the case in which the channels are not fading, ETE switches its channels at half the speed of TE. This effect becomes more remarkable with the growth of the amount of available channels, that is, when the selected configuration is optimal with higher probability.

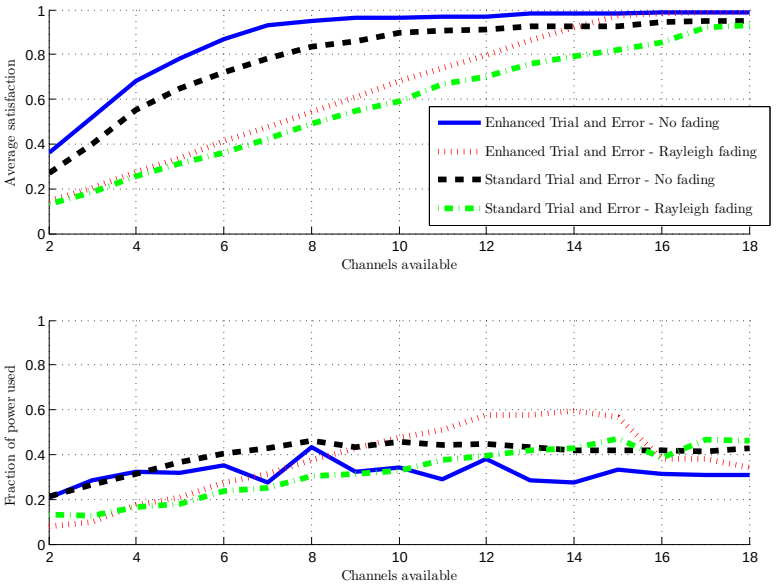

FIGURE 4.19: Comparison between TE and ETE performances in a static scenario, both with Rayleigh fading channels and with block fading ones. The upper plot reports TE and ETE's AS, the lower plot reports TE and ETE's APC.

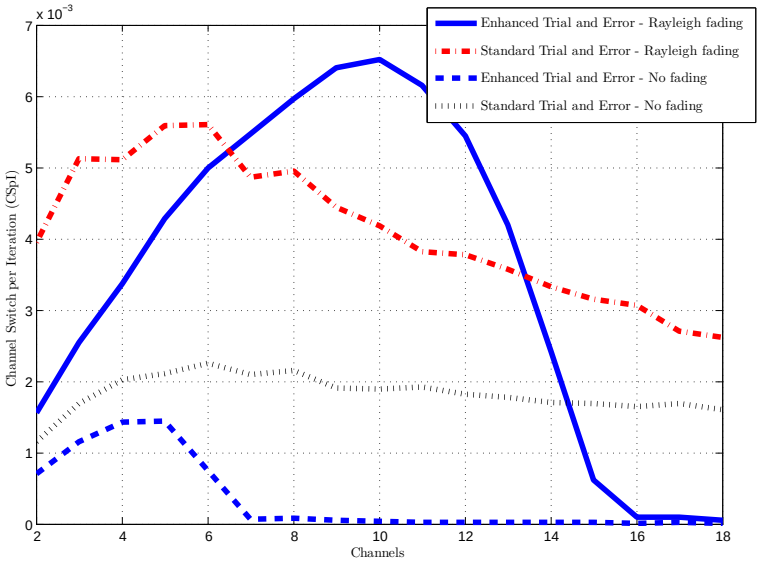

FIGURE 4.20: Channel switch per iteration as a function of the available channels for both ETE and TE learning algorithms in fading and non-fading environment.

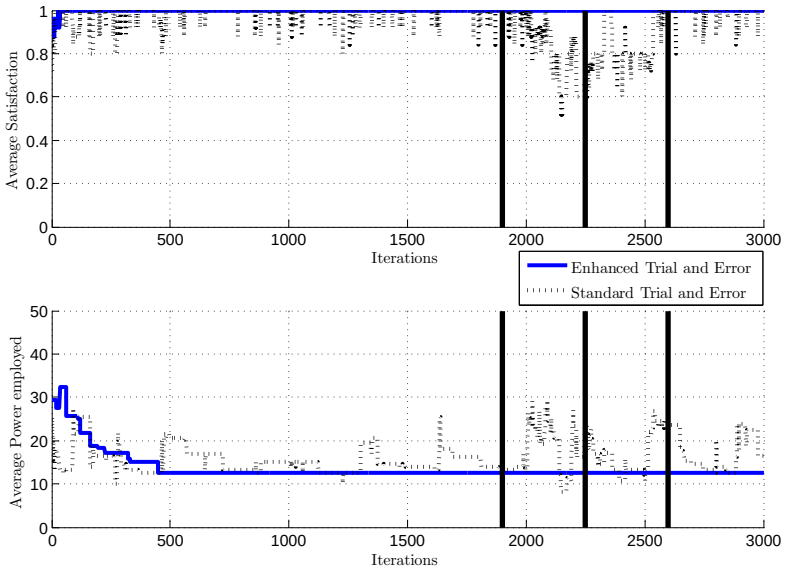

FIGURE 4.21: Achieved AS and APC for both TE and ETE. The black vertical lines indicate, approximately, the moment when the moving cluster is close enough to create interference to the others ($t = 1875$), when it is aligned with the others ($t = 2250$) and when it begins to be far enough ($t = 2575$).

In the case in which the channels are fading, instead, the amount of changes in ETE increases. This is due to the fact that to achieve a well performing working point, it is necessary to jump often from a channel to another. In other words, it is necessary to follow the variations of the channels and consequent variation of the optimal configuration.

4.5.3.2 Mobile DSCN

In this section, ETE and TE performance are compared under the mobile DSCN depicted in Figure 3.2. Assume $K = 4$ clusters to be aligned and sharing the spectrum while a fifth cluster is far away enough to be creating little interference.

Figure 4.21 reports the AS and APC for both TE and ETE as a function of the algorithm's iterations. ETE performs better than of TE in both metrics. This is due to the effect of the stabilization that, when a configuration is well performing, reduces the experimentations. This fact is also sustained by the results reported in Figure 4.22 and Figure 4.13. Both figures report the channel chosen by the CHs (each channel is represented by a different color) and the power used in the cluster for the transmissions, represented by the dimension of the line.

Figure 4.22 shows that very few iterations are sufficient in order to achieve an optimal configuration, and, once achieved, it stays stable for the whole duration of the experiment. In the meanwhile, the power level decreases in order to save energy. Since

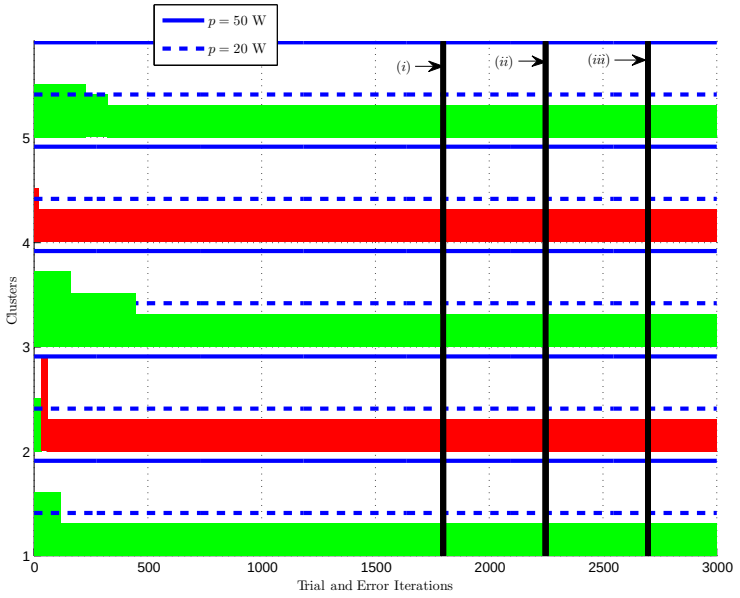

FIGURE 4.22: Progress of the channels for the ETE learning algorithm in a mobility scenario. To each color corresponds a channel, the dimension of the line represents the power level chosen for the communication. On the y-axis, the cluster in mobility is represented by the number 3.

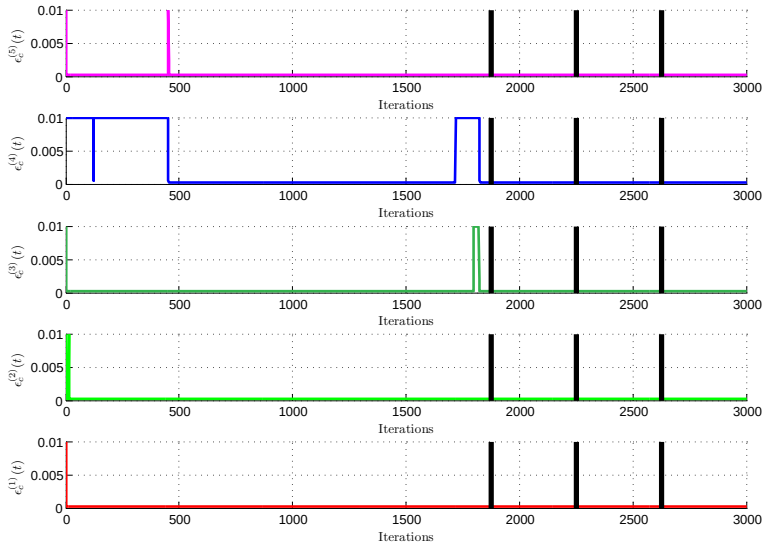

FIGURE 4.23: Value of $\epsilon_c^{(k)}(t)$ versus ETE's iterations in the mobility scenario.

the level of satisfaction in the clusters is high, following (4.11) the value of $\epsilon_c^{(k)}$ decreases and the configuration does not change, as depicted in Figure 4.23

4.5.4 Performance Evaluation and comparisons

This section shows the gain due to the enhancement of the algorithm and compares it with several existing ones such as the greedy based decentralize control algorithm

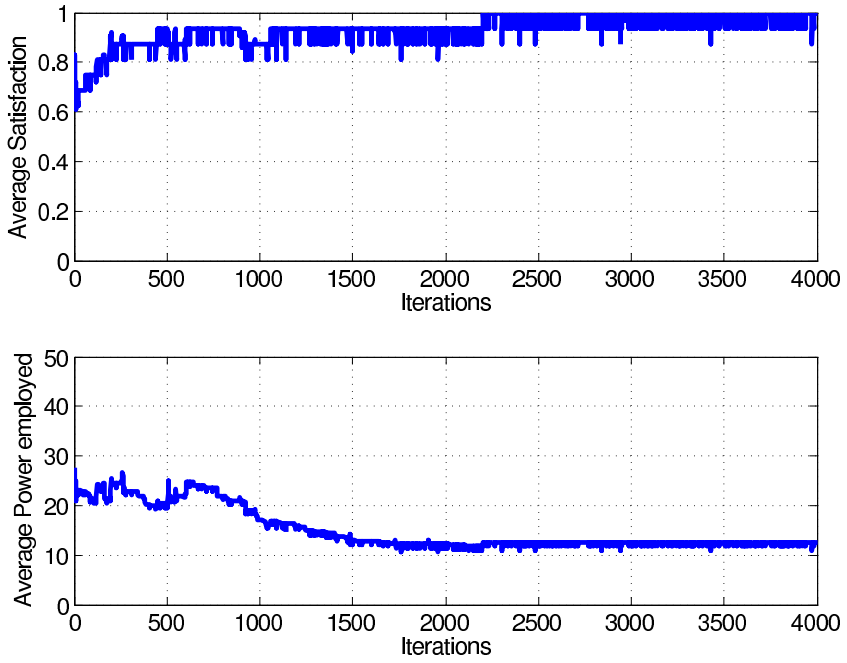

FIGURE 4.24: The upper plot represents the AS, the lower plot represents the APC.

(GBDCA) [43] and the IWF [79].

The simulation scenario is represented in Figure 3.1. Consider a static network composed of $K = 16$ clusters each with $N_k = 4$ links, $C = 10$ channels, and the maximum power $P_{\max} = 50$W is quantized in $Q = 8$ logarithmic levels. The results are reported in Figure 4.24, where the upper plot represents the AS while the lower plot shows the APC. The figure shows that the TE algorithm is able to drive the network to an almost full satisfaction by employing on average only 10W. Note that, even though the first visit to an NE may happen quite late, the global performance at non-equilibrium states is high. This is due to the fact that the probability of playing an action grows with the social welfare of the action itself [101]. Second, Figure 4.24 shows that even when an equilibrium is achieved, the system sometimes attempts to use sub-optimal action profiles. This is due to the stochastic nature of the TE learning algorithm. Note that there exists a natural tradeoff between the time needed to visit an NE and stability of such an equilibrium. In order to decrease the time needed to visit an NE, the experimentation probability needs to be large while, in order to improve the stability it needs to be small.

Furthermore, the TE learning algorithm is compared with the GBDCA described in [43]. Briefly, this algorithm solves the graph-coloring problem, by letting each CC detect the channel employed by its neighbors. If a CC detects that it is using a channel already occupied by one of its neighbors then it chooses randomly another channel among the free ones. If no channel is free, then it does not change its strategy. Since this algorithm does not consider a power allocation policy, its transmission power is set to $P_{\max}$. In this context, the GBDCA is compared with the TE learning algorithm when the quantization

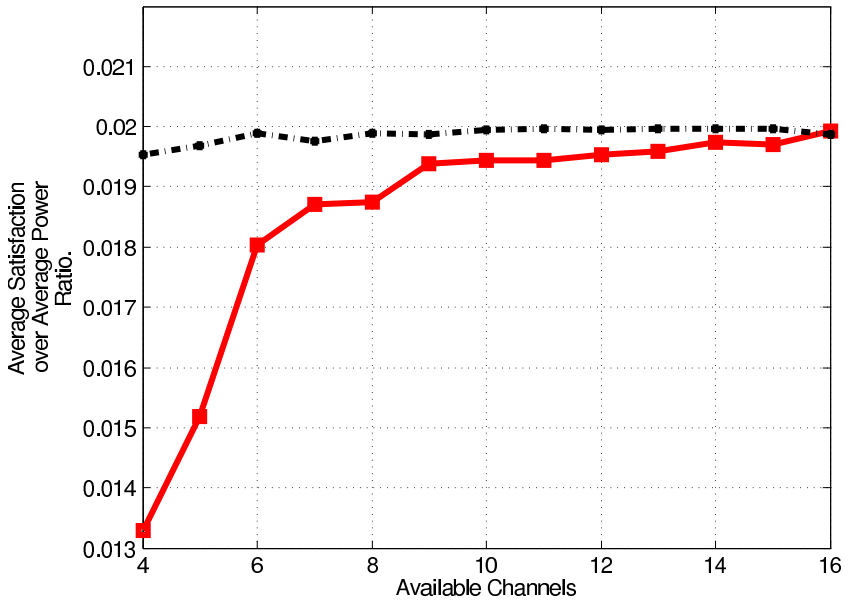

FIGURE 4.25: Performance comparison between TE and the GBDCA in terms of average number of constraints satisfied over average used power. The dashed line is the performance of TE and the continuous line the one of GBDCA .

levels are reduced to $Q = 2$, i.e., an ON-OFF policy. The results, in terms of the ratio $\frac{\sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{\xi_\ell > \Gamma\}}}{\sum_{k \in \mathcal{K}} p_k}$ are reported in Figure 4.25. The TE learning algorithm allows the clusters that cannot satisfy their constraints to stop the transmission for a short period of time, which increases the efficiency.

The following compares the performance of the TE learning algorithm with the one of synchronous IWF and the global optimum. Consider $K = 16$ clusters, $N_k = 1$ link per cluster, $C = 5$ channels, $Q = 5$ power levels and a target SINR $\Gamma = 10$ dB. In the synchronous IWF each transmitter has full knowledge of the transmit channel state information; each transmitter may exploit multiple channels; the power allocation routine happens at the same instant for all transmitters; and each transmitter attempts to achieve a transmission rate equal to $\log_2(1 + \Gamma)$ with the minimum necessary power. The results of the experiment are reported in Figure 4.26.

The first figure reports the AS in the upper plot and the APC in the lower plot. In these plots, the dashed line represents the global optimum, the continuous red line the performance of TE algorithm and the dotted line the performance of the synchronous IWF. The action profiles chosen by the TE algorithm approach the global optimum both in terms of constraints satisfaction and in terms of power drain. The synchronous IWF , even though it is allowed to exploit a larger amount of information, is not able to select an action that satisfies the constraints for a large proportion of the links.

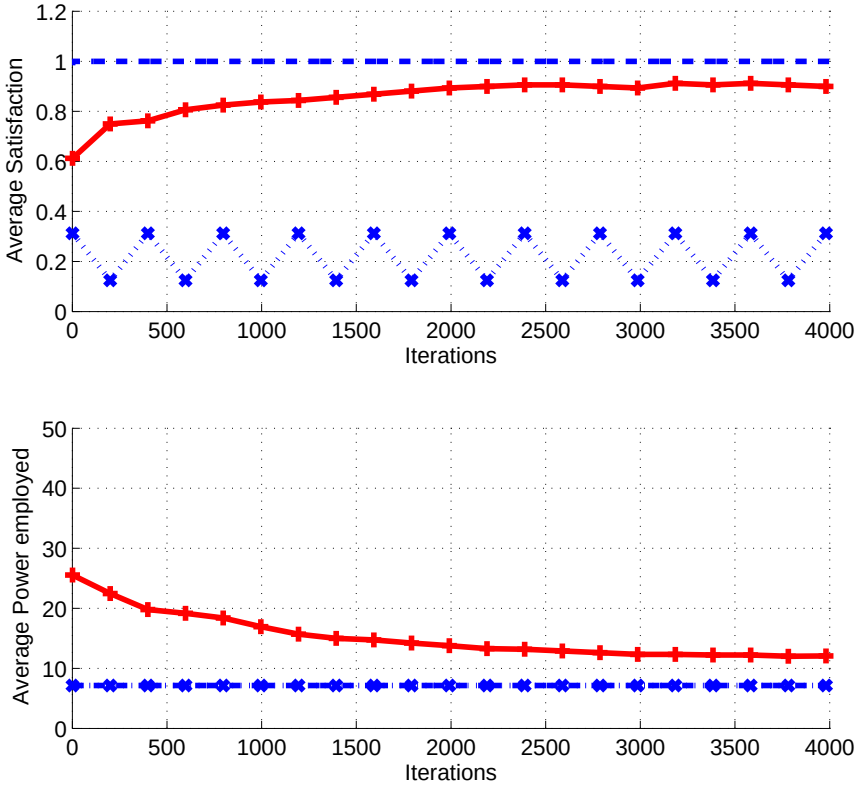

FIGURE 4.26: Performance comparison between TE (red continuous line), the synchronous IWF (dotted line) and the global optimum (dashed line). We represent, in the upper plot, the AS, in the lower plot, the APC . We run 400 iterations of each algorithm on a network composed of $K = 16$ clusters, each populated with one link, $C = 5$ channels, $Q = 5$ power levels, a maximum available power of $P_{\max} = 50$W.

4.5.5 High Fidelity Simulator

In order to allow the evaluation of learning algorithms at operational level, the project CORASMA developed a HiFi encompassing the three first layers of the international standard organization (ISO) model, namely the physical (PHY) layer, media access control (MAC) layer and the network (NET) layer [20]. In this context, HiFi means that the detail level of implementation is enough to replicate behavior of real systems for these layers. At PHY layer, channel coding and decoding is implemented as well as modulation and demodulation in baseband. Transmitted signals are sent through a propagation channel that integrates a digital terrain model including the above ground such as buildings. At the MAC layer, all the protocols are implemented including the signaling messages needed to operate the protocols. This is of paramount importance when evaluating learning solutions since it allows to assess the extra signaling required as well as their sensitivity to the loss of signaling. At the NET layer true implementation of routing protocols is done transmitting internet protocol (IP) datagrams through the network. The interest in implementing such detail stems from the fact that it allows to capture the impact of the lower layers behavior on datagrams and IP signaling.

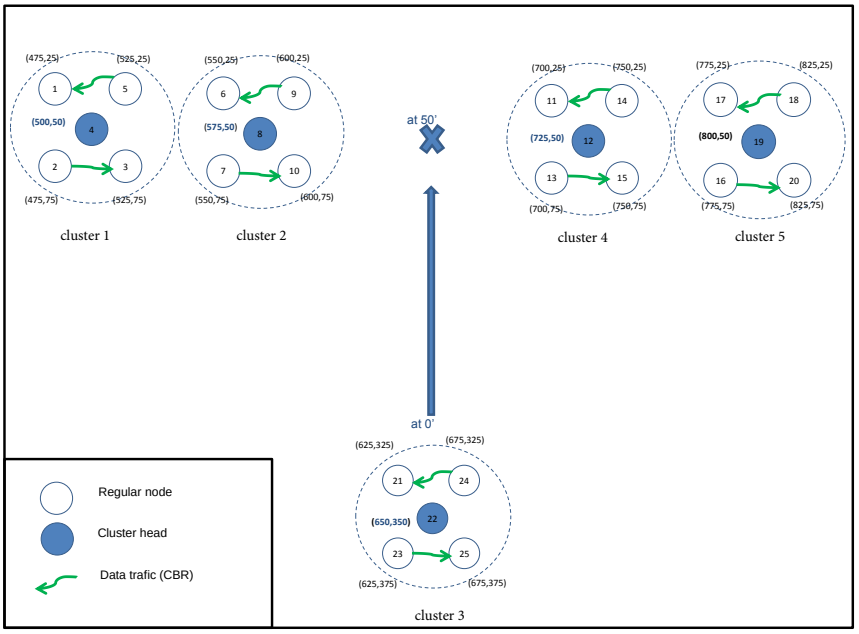

FIGURE 4.27: Scenario description used in the CORASMA simulator (distances in meter).

4.5.5.1 Simulation Results

This subsection reports the preliminary results from the CORASMA simulator on the ETE applied to DSCNs.

The experiment implements the scenario depicted in Figure 4.27, which is the CORASMA equivalent of the mobile DSCN introduced in Section 3.1. The purpose is to assess whether realistic constraints can reduce the ETE learning algorithm performance or the behavior of the ETE impacts the upper layers. Furthermore, a comparison with and it is also possible to compare the frequency channels set by TE with the one of GBDCA [43]. In this scenario, a user datagram protocol (UDP) constant bit rate traffic of 6500 bytes/s is implemented between nodes and indicated inside the figures with green arrows.

In order to analyze and compare the performance between GBDCA and ETE, we make use of the statistical metric display tool developed with the CORASMA simulator. Note that in the following figures, the red curves and blue curves represent the performance of GBDCA and the ETE respectively. Figure 4.28 provides the frequency selection for each CH along time for both the GBDCA and ETE.

Remarkably, GBDCA does not change the channel-cluster association during the simulation whereas the ETE solution does. In particular, one can observe that the frequency selection starts varying around 40 seconds. This is due to the fact that the mobile cluster becomes close to the other clusters and starts interfering. After 55 seconds

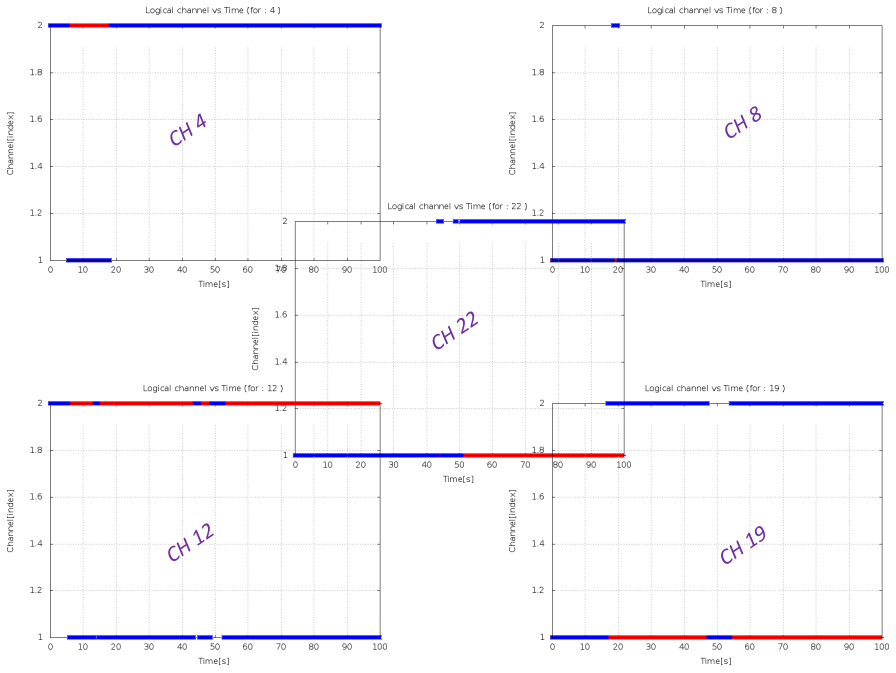

FIGURE 4.28: Five clusters' logical channel along time for the two solutions, BW (static frequency) and TE - Only two available logical channels.

(i.e., 15 seconds for convergence), the channel selection stabilizes. Therefore, in the GBDCA case, clusters 2 and 5 with interfere with each other reducing the respective performance. This illustrates one drawback of the GBDCA algorithm that is not based upon interference measurements, rather on an heuristic collaboration among the CHs and that can remain stuck in an interfering configuration without the capacity to resolve it.

This reflects at the IP layer as reported in Figure 4.29. This figure reports the throughput achieved in the communication between two devices belonging to cluster 2. The plot shows that the ETE succeeded to adapt the channel and power such that the selected configuration does not create significant interference. The GBDCA throughput drops after around 40 seconds due to the channel-cluster association, while the one selected by the ETE learning algorithm fast reacts to the arising interference.

4.6 Closing Remarks

In this chapter strong connections between the solutions to a centralized network optimization problem and the Nash equilibria of a given game has been established via the design of the corresponding utility functions. More specifically, it has been proven that by properly choosing the utility function, it is possible to make a decentralized network to be stable at a global optimal operating point. More importantly, it has been shown that such equilibria can also be achieved by using learning algorithms following

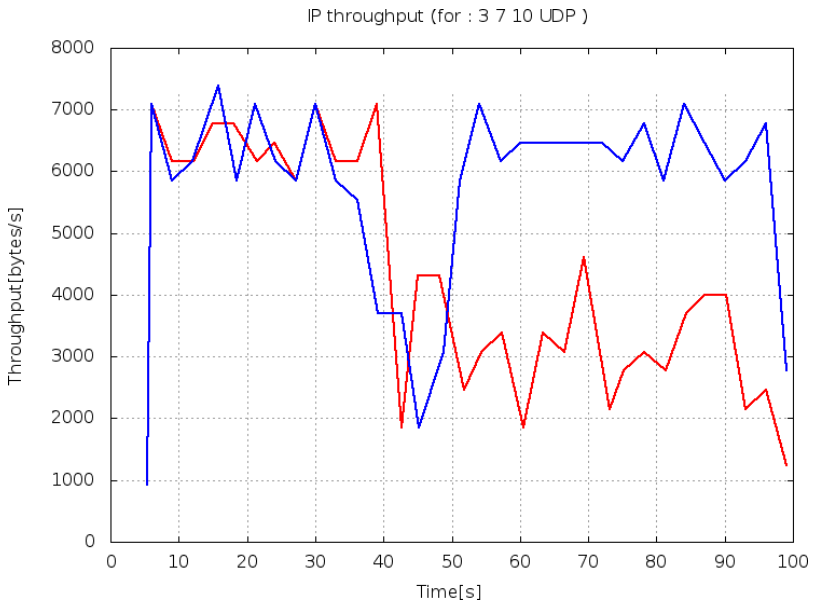

FIGURE 4.29: IP throughput at node 10 received from node 7.

the paradigm of trial and error. The converging capabilities of such an algorithm are studied theoretically, and an approximated close form expression of the converging time and the average time spent on the equilibrium is given and validated numerically.

The performance of the various learning algorithm presented in Chapter 2, namely the best response dynamics, the fictitious play, the smooth fictitious play, the regret matching, the reinforcement learning, the joint utility and strategy estimation based reinforcement learning and trial and error, has been established through extensive numerical simulations. The main drawbacks of the trial and error learning algorithm have been identified in the drop of performance due to the instability of the cluster-channel association and the sub-optimal experimentation policy. An enhanced version of the algorithm is proposed and shown to reduce the instability and improve the performance both in static and in mobile networks.

Several open issues remain however to be solved. The need for a constant feedback from the links to the cluster head can still pose a performance drop and it is clearly a weak point of the system.

Chapter 5

Conclusions and Outlook

5.1 Conclusions

Next generation telecommunications networks are challenged by the increased demand of wireless connectivity. In order to face the growing demand for wireless data services, new generation mobile communication systems will likely be composed of densely deployed, short ranged base stations, e.g., SCs and indoor femto-cells. These devices are envisioned to be deployed with a minimum of planning, and are thought to be capable of continuously exploring their environment in order to optimally configure their parameters. Networks characterized by the massive presence of such devices have a growing need for exploiting intelligently the limited available resources. As a consequence DSCNs demand efficient algorithms able to optimally configure the network's parameters.

Furthermore, in military and emergency recovering networks, there is a natural need for secrecy and flexibility of communications. State of the art military communications base their managing of the spectrum on a pyramidal hierarchy of man made decisions, and a completely centralized resource management. However, in this context, the presence of fixed telecommunications infrastructures is neither practical nor desirable. The evident constraints due to harshness of conditions sums up with the weak points that the presence of a BS offers to a malicious user making self configuration an even more desirable feature. Hence, mitigating interference, avoiding collisions and reducing the power consumption in such networks is of paramount importance. However, due to the unpredictability of the wireless conditions, self-configuration functionalities become a necessary feature. Such a complex network, composed by intelligent self-configuring devices, requires new theoretical frameworks in order to analyze their performance. Efficient mechanisms capable of configuring the network's transmission parameters in an efficient way, and theoretical means that help in assessing the performance of these configurations are a necessity.

In this thesis, a game theoretical model of DSCNs has been proposed and several learning algorithms for self-configuring networks have been studied and discussed. The pertinence of these algorithms for wireless communications in terms of system constraints (continuous or discrete actions' set, required information, information assumptions, synchronization, signaling, etc.) have been identified, as well as the performance criteria such as the utility achieved at the steady state, convergence speed, etc. A precise link between algorithms and game theoretical relevant equilibrium concepts has been established. This link allows a network designer to create particular actions' set and utility functions in order to let the equililbria have particularly interesting characteristics, such as high fairness, high global performance, etc. The limits and drawbacks of these algorithms have been assessed and a particular algorithm, namely the trial and error learning algorithm, has been selected for configuring a military DSCN. The main reasons is that asymptotic learning algorithms demand particular games' structure in order to converge to the equilibrium. On the other hand, a game model of a DSCN rarely respects any of these types. This makes the TE learning algorithm a suitable candidate, given its capability of converging in different games. This algorithm's peculiarities can summarized as follows:

- It is composed of a state machine running at each decision taker;
- It requires minimum knowledge on the game played;
- It requires only a numerical estimation of the utility at each iteration;
- Its stochastically stable states, i.e. the states that are most likely played in the long run, are the Nash equilibria with highest social welfare;
- It does need reinitialization and it responds quickly to changes in the network, thanks to the absence of an asymptotic convergence state.

The main problem associated to TE learning algorithm has been identified in the instability of the channel-cluster association, and in the sub-optimal configuration selection policy. To amend these issues an enhancement version of the algorithm has been developed. Its main characteristics are based on the presence of double experimentation factor, and smarter configuration policy that tests only possibly optimal configurations. The role of each new parameter has been discussed and their effect in the converging capabilities of the algorithm has been assessed. In particular, has been noticed that small experimentation frequency in the channel-cluster association are recommended in static networks, while the presence of unpredictable events such as fading, increase the need for fast response. The capabilities of this algorithms are then assessed under various scenarios, both static and mobile in presence and absence of fading channels.

Even though the enhanced trial and error algorithm has shown remarkable capabilities in configuring DSCNs, some problems remains unsolved. In real networks,

unpredictable events may also come from inside the cluster. Devices that wish not to transmit, radios that are no longer functional, an erroneous evaluation of the perceived QoS or even corrupted feedback can deteriorate the reliability of the CH's estimation of the utility function.

5.2 Outlook

The work conducted in this thesis leads to several interesting possible expansions described hereunder.

5.2.1 Decentralized Self-configuring Networks' Modeling

GT has proven as a powerful tool to model DSCNs. The growing list of mathematical refinements of the theory such as stochastic games are attempts to better model behaviors of autonomous player in any real environment. However, even this type of games fails in perfectly modeling a DSCN. For instance, positioning of the devices and their appearance or disappearance are hardly modeled. A relevant contribution to solving this problem may come from stochastic geometry, a mathematical framework that develops network models in which the locations of the devices, and structure of the network are random variables. A complete mathematical characterization of DSCNs could improve the understanding of its mechanism and lead to better self-configuring algorithms.

5.2.2 Algorithms Design

As already mentioned, several algorithmic approaches, alternative to learning algorithms, exist in the literature. An iterative procedure that is able to learn a particular equilibrium or that shows a predictable and evaluable outcome in a wide range of cases is however missing. There is a growing interest in the use of variational inequalities in order to achieve predictable steady states in a wide set of scenario's typologies. Moreover, algorithms that can better absorb information from sensing and maybe even triggering the sensing of particular parameters are a viable and interesting path to improve the quality of configurations.

5.2.3 Game Theory

We believe that the role of GT is far from being finished in the field of DSCNs. As it was shown, some level of centralization, even though just local centralization still persists in real networks. A possible development in this sense is given by the idea of studying the local centralization through cooperative GT. This could lead to more efficient and

dynamic CH selection algorithms or to a decentralized control algorithm that enables the network to obtain the same result of the centralized one without the need for a central controller.

List of Figures

List of Tables

[1] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[2] J. Hoydis, M. Debbah *et al.*, "Green, cost-effective, flexible, small cell networks," *IEEE Communications Society MMTC*, vol. 5, no. 5, pp. 23–26, 2010.

[3] S. Landström, A. Furuskär, K. Johansson, L. Falconetti, and F. Kronestedt, "Heterogeneous networks–increasing cellular capacity," *The data boom: opportunities and challenges*, p. 4, 2011.

[4] M. Debbah, "Mobile flexible networks: The challenges ahead," in *International Conference onAdvanced Technologies for Communications (ATC)*. IEEE, 2008, pp. 3–7.

[5] W. Kiess and M. Mauve, "A survey on real-world implementations of mobile ad-hoc networks," *Ad Hoc Networks*, vol. 5, no. 3, pp. 324–339, Apr. 2007.

[6] J. Mitola and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[7] L. Rose, S. Lasaulce, S. M. Perlaza, and M. Debbah, "Learning equilibria with partial information in decentralized wireless networks," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 136–142, Aug. 2011.

[8] I. Macaluso, L. D. Silva, and L. Doyle, "Learning Nash equilibria in distributed channel selection for frequency-agile radios," in *Proc. Workshop on Artificial Intelligence for Telecommunications and Sensors Networks (WAITS)*, Montpellier, France, Apr. 2012.

[9] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Communications Surveys Tutorials*, vol. 1, no. 99, pp. 1–25, Apr. 2013.

[10] D. Gesbert, S. Kiani, A. Gjendemsjo, and G. ien, "Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks," *Proc. of the IEEE*, vol. 95, no. 12, pp. 2393–2409, Dec. 2007.

[11] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 2, pp. 180–194, Apr. 2012.

[12] C. Tekin, M. Liu, R. Southwell, J. Huang, and S. Ahmad, "Atomic congestion games on graphs and their applications in networking," *IEEE ACM Transactions on Networking*, vol. 20, no. 5, pp. 1541–1552, Oct. 2012.

[13] R. D. Yates, D. Tse, and Z. Li, "Secret communication on interference channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Toronto, Canada, Jul. 2008.

[14] R. Berry and D. Tse, "Shannon meets nash on the interference channel," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2821–2836, May 2011.

[15] S. M. Perlaza, R. Tandon, H. V. Poor, and Z. Han, "The Nash equilibrium region of the linear deterministic interference channel with feedback," in *Proc. 50th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Oct. 2012.

[16] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[17] A. Ghasemi and E. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, 2008.

[18] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 1, Feb. 2004, pp. 772–776.

[19] Combined Communications and Electronics Board, "Guide to spectrum management in military operations, acp 190(c)," Sep. 2007.

[20] L. Rose, R. Massin, L. Vijayandran, M. Debbah, and C. J. Le Martret, "CORASMA program on cognitive radio for tactical networks: High fidelity simulator and first results on dynamic frequency allocation," in *in Proc. of the IEEE Military Communication Conference, (Milcom)*, San Diego,CA, USA, 2013.

[21] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Journal of Computer networks (Elsevier)*, vol. 50, pp. 2127–2159, 2006.

[22] B. Scheers, A. Mahoney, and H. Akermark, "A realistic roadmap for the introduction of dynamic spectrum management in military tactical radio communication," in *Military Communications and Information Systems Conference (MCC)*, 2012, pp. 1–8.

[23] T. Hong and Z. Jie, "A framework of intelligent decision support system of military communication network effectiveness evaluation," in *Proc. of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 4, 2008, pp. 518–521.

[24] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, 2008.

[25] T. C. Clancy III, "Dynamic spectrum access in cognitive radio networks," Ph.D. dissertation, University of Maryland, 2006.

[26] D. N. Hossain, Ekram and Z. Han. Cambridge University Press, 2009.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[28] B. Johansson, "On distributed optimization in networked systems," Ph.D. dissertation, KTH, 2008.

[29] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 24, no. 8, pp. 1439–1451, 2006.

[30] D. Mosk-Aoyama, T. Roughgarden, and D. Shah, "Fully distributed algorithms for convex optimization problems," *DISC*, pp. 492–493, 2007.

[31] Y. Shi and Y. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," in *Proc. of The IEEE 27th Conference on Computer Communications (INFOCOM)*, Apr. 2008, pp. 1292–1300.

[32] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *in Proc. of the International Conference on Computer Communications (INFOCOM)*, Miami, FL, USA, Mar. 2005, pp. 1653–1664.

[33] J. Holland, *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* University of Michigan Press, 1975.

[34] I. Rechenberg, "Evolutionsstrategie–optimierung technisher systeme nach prinzipien der biologischen evolution," 1973.

[35] K. A. De Jong and J. Sarma, "On decentralizing selection algorithms." in *ICGA*, vol. 95, 1995, pp. 17–23.

[36] L. Boumediene, Z. Gao, S. Liu, S. Leghmizi, and R. Yang, "Genetic algorithm-based approach to spectrum allocation and power control with constraints in cognitive radio networks," *Research Journal of Applied Sciences*, vol. 5, 2013.

[37] T. W. Rondeau, B. Le, C. J. Rieser, and C. W. Bostian, "Cognitive radios with genetic algorithms: Intelligent control of software defined radios," in *Software defined radio forum technical conference*, Apr. 2004, pp. C3–C8.

[38] C. Rieser, T. Rondeau, C. Bostian, and T. Gallagher, "Cognitive radio testbed: further details and testing of a distributed genetic algorithm based cognitive engine for programmable radios," in *IEEE Military Communications Conference (MILCOM)*, vol. 3, Oct. 2004, pp. 1437–1443.

[39] C. Doerr, D. Sicker, and D. Grunwald, "Experiences implementing cognitive radio control algorithms," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, Jun. 2007, pp. 4045–4050.

[40] D. B. West *et al.*, *Introduction to graph theory*. Prentice hall Englewood Cliffs, 2001, vol. 2.

[41] S. Dickinson, M. Pelillo, and R. Zabih, "Introduction to the special section on graph algorithms in computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1049–1052, 2001.

[42] C. L. Baber, "An introduction to list colorings of graphs," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2009.

[43] T.-C. Hou and T.-J. Tsai, "On the cluster based dynamic channel assignment for multihop ad hoc networks," *Journal of Communications and Networks*, vol. 4, no. 1, pp. 40–47, Mar. 2002.

[44] J. Wang, Y. Huang, and H. Jiang, "Improved algorithm of spectrum allocation based on graph coloring model in cognitive radio," in *Proc. of WRI International Conference on Communications and Mobile Computing, 2009. CMC'09.*, vol. 3, Sep. 2009, pp. 353–357.

[45] Y. Chen, N. Han, S. Shon, and J. M. Kim, "Dynamic frequency allocation based on graph coloring and local bargaining for multi-cell wran system," in *Proc. of Conference on Communications, 2006. APCC '06*, vol. 1, May 2006, pp. 1–5.

[46] D. J. Leith, P. Clifford, V. Badarla, and D. Malone, "Wlan channel selection without communication," *Computer Networks*, vol. 56, no. 4, pp. 1424–1441, 2012.

[47] P. Clifford and D. J. Leith, "Channel dependent interference and decentralized colouring," in *Network Control and Optimization*. Springer, 2007, pp. 95–104.

[48] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[49] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *Machine Learning: ECML 2005*. Springer, 2005, pp. 437–448.

[50] J. C. Gittins, "Bandit processes and dynamic allocation indices," *Series B (Methodological) Journal of the Royal Statistical Society.*, pp. 148–177, 1979.

[51] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multi-armed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1902–1916, 2013.

[52] A. Anandkumar, N. Michael, A. K. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 731–745, 2011.

[53] K. Liu and Q. Zhao, "A restless bandit formulation of opportunistic access: Indexablity and index policy," in *Proc. of 5th IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops (SECON)*. IEEE, 2008, pp. 1–5.

[54] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation," in *Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum*. IEEE, 2010, pp. 1–9.

[55] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: learning under competition," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2010, pp. 1–9.

[56] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5667–5681, 2010.

[57] ——, "Distributed learning in multi-armed bandit with multiple players," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.

[58] M. N. Katehakis and A. F. Veinott, "The multi-armed bandit problem: decomposition and computation," *Mathematics of Operations Research*, vol. 12, no. 2, pp. 262–268, 1987.

[59] J. V. Neumann and O. Morgenstern, "Theory of games and economic behavior," *Princeton University Press*, 1944.

[60] M. J. Osborne, *An introduction to game theory*. Oxford University Press New York, 2004, vol. 3, no. 3.

[61] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, "Coalitional game theory for communication networks," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 77–97, 2009.

[62] J. F. Nash, "Equilibrium points in n-person games," *in Proc. of the National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, 1950.

[63] S. M. Perlaza, E. V. Belmega, S. Lasaulce, and M. Debbah, "On the base station selection and base station sharing in self-configuring networks," *3rd ICST/ACM International Workshop on Game Theory in Communication Networks*, Oct.

[64] E. Belmega, S. Lasaulce, and M. Debbah, "Decentralized handovers in cellular networks with cognitive terminals," *3rd Intl. Symp. on Communications, Control and Signal Processing - ISCCSP*, March 2008.

[65] E. G. Larsson, E. A. Jorswieck, J. Lindblom, and R. Mochaourab, "Game Theory and the Flat-Fading Gaussian Interference Channel: Analyzing Resource Conflicts in Wireless Networks," *IEEE signal processing magazine (Print)*, vol. 26, no. 5, pp. 18–27, 2009.

[66] R. Mochaourab and E. Jorswieck, "Walrasian equilibrium in two-user multiple-input single-output interference channels," in *Proc. IEEE International Conference on Communications Workshops (ICC),*, Jun. 2011, pp. 1–5.

[67] L. Rose, S. M. Perlaza, and M. Debbah, in *Proc. of IEEE Workshop on Game Theory and Resource Allocation for 4G*, Kyoto, Japan, Jun., pp. 1–6.

[68] S. M. Perlaza, "Game theoretic approaches to spectrum sharing in decentralized self-configuring networks," Ph.D. dissertation, Télécom ParisTech, Jul. 2011.

[69] E. Jorswieck and R. Mochaourab, "Power control game in protected and shared bands: Manipulability of nash equilibrium," in *International Conference on Game Theory for Networks, 2009. GameNets '09.*, 2009, pp. 428–437.

[70] T. Alpcan, T. Başar, R. Srikant, and E. Altman, "Cdma uplink power control as a noncooperative game," *Wireless Networks*, vol. 8, no. 6, pp. 659–670, 2002.

[71] G. Bacci, M. Luise, H. Poor, and A. Tulino, "Energy efficient power control in impulse radio uwb wireless networks," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 3, pp. 508–520, 2007.

[72] Z. Han, Z. Ji, and K. Liu, "Fair multiuser channel allocation for ofdma networks using nash bargaining solutions and coalitions," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, 2005.

[73] D. Niyato and E. Hossain, "Cognitive radio for next-generation wireless networks: An approach to opportunistic channel selection in IEEE 802.11-based wireless mesh," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 46–54, Feb. 2009.

[74] D. Fudenberg and J. Tirole, *Game Theory.* Cambridge, MA, USA: The MIT Press, 1991.

[75] J. S. Jordan, "Three problems in learning mixed-strategy nash equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 368–386, 1993.

[76] V. Borkar, "Reinforcement learning in markovian evolutionary games," *Advances in Complex Systems (ACS)*, vol. 05, no. 01, pp. 55–72, 2002.

[77] H. P. Young, *Strategic Learning and Its Limits (Arne Ryde Memorial Lectures Sereis)*. Oxford University Press, USA, 2004.

[78] L. Rose, S. M. Perlaza, M. Debbah, and C. J. Le Martret, "Distributed power allocation with SINR constraints using trial and error learning," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 1835–1840.

[79] G. Scutari, D. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory – part II: Algorithms," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1250–1267, Mar. 2008.

[80] P. Sastry, V. Phansalkar, and M. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 5, pp. 769–777, May 1994.

[81] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 7, pp. 3202–3212, 2013.

[82] J.-S. P. G. Scutari, D. P. Palomar and F. Facchinei, "Flexible design of cognitive radio wireless systems: From game theory to variational inequality theory," *IEEE Signal Processing Mag.,*, vol. 26, no. 5, pp. 107–123, Sept. 2009.

[83] A. Leshem and E. Zehavi, "Game theory and the frequency selective interference channel - a tutorial," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 28–40, 2009.

[84] S. Vishwanath, W. Rhee, N. Jindal, S. Jafar, and A. Goldsmith, "Sum power iterative waterfilling for Gaussian vector broadcast channels," in *Proc. of IEEE International symposium on Information theory (ISIT)*, June 2003, pp. 467–470.

[85] J.-S. Pang, G. Scutari, D. P. Palomar, and F. Facchinei, "Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3251–3271, Jun. 2010.

[86] V. L. Nir and B. Scheers, "Improved coexistence between multiple cognitive tactical radio networks by an expert rule based on sub-channel selection," in *Proc. of Wireless Innovation Forum European Conference on Communications Technologies and Software Defined Radio SDR'11 (WInnComm-Europe)*, Brussels, Belgium, Jun. 2011.

[87] V. Le Nir and B. Scheers, "Autonomous dynamic spectrum management for co-existence of multiple cognitive tactical radio networks," in *Proc. of IEEE Fifth International Conference on Cognitive Radio Oriented Wireless Networks Communications (CROWNCOM)*, Cannes, France, Jun. 2010.

[88] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. on Info. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[89] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1991.

[90] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, "Distributed power allocation with rate constraints in Gaussian parallel interference channels," *IEEE Trans. on Info. Theory*, vol. 54, no. 8, pp. 3471–3489, Aug. 2008.

[91] G. Scutari, D. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory – Part I: Nash equilibria," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1230–1249, Mar. 2008.

[92] O. Popescu and C. Rose, "Water filling may not good neighbors make," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, San Francisco, CA, USA, Dec. 2003.

[93] E. Altman, V. Kamble, and H. Kameda, "A Braess type paradox in power control over interference channels," in *6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops (WiOPT)*, Berlin, Germany, May 2008, pp. 555–559.

[94] R. Laraki, "Variational inequalities, system of functional equations, and incomplete information repeated games," *SIAM Journal on control and optimization*, vol. 40, no. 2, pp. 516–524, 2001.

[95] P. T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications," *Mathematical programming*, vol. 48, no. 1-3, pp. 161–220, 1990.

[96] G. J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*. Prentice Hall New Jersey, 1995.

[97] H.-S. T. Le, H. D. Ly, and Q. Liang, "Opportunistic spectrum access using fuzzy logic for cognitive radio networks," *International Journal of Wireless Information Networks*, vol. 18, no. 3, pp. 171–178, 2011.

[98] B. Babadi and V. Tarokh, "GADIA: A greedy asynchronous distributed interference avoidance algorithm," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6228–6252, Dec. 2010.

[99] L. Iacobelli, F. Scoubart, D. Pirez, P. Fouillot, R. Massin, C. Lefebvre, C. Le Martret, and V. Conan, "Dynamic frequency allocation in ad hoc networks," in *Proc. of 2nd International Workshop on Cognitive Information Processing (CIP)*, Feb. 2010, pp. 145–150.

[100] H. P. Young, "Learning by trial and error," University of Oxford, Department of Economics, Economics Series Working Papers 384, 2008.

[101] B. S. Pradelski and H. P. Young, "Learning efficient Nash equilibria in distributed systems," Tech. Rep., Sep. 2010.

[102] V. Srivastava, J. Neel, A. MacKenzie, R. Menon, L. Dasilva, J. Hicks, J. Reed, and R. Gilles, "Using game theory to analyze wireless ad hoc networks," *IEEE, Communications Surveys Tutorials*, vol. 7, no. 4, pp. 46–56, 2005.

[103] A. MacKenzie and S. Wicker, "Game theory and the design of self-configuring, adaptive wireless networks," *IEEE Communications Magazine*, vol. 39, no. 11, pp. 126–131, Nov 2001.

[104] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 23, no. 2, pp. 201–220, 2005.

[105] E. Altman and Z. Altman, "S-modular games and power control in wireless networkse," *IEEE Trans. Automat. Contr.s*, vol. 48, no. 5, pp. 839–842, May 2003.

[106] G. Bacci, L. Sanguinetti, M. Luise, and H. V. Poor, "A game-theoretic approach for energy-efficient contention-based synchronization in ofdma systems," 2013.

[107] Z. Ji and K. Liu, "Cognitive radios for dynamic spectrum access - dynamic spectrum sharing: A game theoretical overview," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 88–94, 2007.

[108] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine, Special issue on Game Theory for Signal Processing*, vol. 26, no. 5, pp. 41–52, Sep. 2009.

[109] D. Braess, *Unternehmensforschung*, vol. 24, no. 5, pp. 258–268, May 1969.

[110] E. Ostrom, "Tragedy of the commons," *The New Palgrave Dictionary of Economics*, 2008.

[111] H. P. Young, "Learning by trial and error," *Games and Economic Behavior*, vol. 65, no. 2, pp. 626–643, Mar. 2009.

[112] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, ser. MIT Press Books. The MIT Press, 1998, vol. 1.

[113] G. Scutari, D. Palomar, and S. Barbarossa, "Simultaneous iterative water-filling for Gaussian frequency-selective interference channels," in *Information Theory, 2006 IEEE International Symposium on*, July 2006, pp. 600–604.

[114] G. W. Brown, "Iterative solution of games by fictitious play," *Activity Analysis of Production and Allocation*, vol. 13, no. 1, pp. 374–376, 1951.

[115] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Int. J. Economic Theory*, vol. 68, pp. 258–265, 1996.

[116] S. Hart and A. Mas-Colell, "A general class of adaptive strategies," *Journal of Economic Theory*, vol. 98, no. 1, pp. 26–54, 2001.

[117] J. R. Marden, G. Arslan, and J. S. Shamma, "Regret based dynamics: convergence in weakly acyclic games," in *Proc. of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007, pp. 194–201.

[118] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *MIT Press, Cambridge, MA*, 1998.

[119] S. M. Perlaza, H. Tembine, and S. Lasaulce, "How can ignorant but patient cognitive terminals learn their strategy and utility?" in *the 11th IEEE Intl. Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010)*, Marrakech, Morocco, June 2010.

[120] D. Fudenberg and D. K. Levine, "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control*, vol. 19, no. 5, pp. 1065–1089, 1995.

[121] J. R. Marden, H. P. Young, and L. Y. Pao", ""achieving pareto optimality through distributed learning"," "2011", "under submission".

[122] H. Wu, Z. Zhong, and L. Hanzo, "A cluster-head selection and update algorithm for ad hoc networks," in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM)*, Feb. 2010, pp. 1–5.

[123] S. Koslowski, J. Elsner, F. K. Jondral, S. Couturier, C. Keip, and O. Bettinger, "Distributed localized interference avoidance for dynamic frequency hopping ad hoc networks."

[124] L. Rose, S. M. Perlaza, C. J. Le Martret, and M. Debbah, "Achieving pareto optimal equilibria in energy efficient clustered ad hoc networks," in *in Proc. of IEEE International Conference on Communications (ICC)*, Budapest, Hungary, June 2013, pp. 1491–1495.

[125] ——, "Self-organization in decentralized networks: A trial and error learning approach," *to appear in IEEE Transactions on Wireless Communications*, 2013.

[126] L. Rose, C. J. Le Martret, and M. Debbah, "Channel and power allocation algorithms for ad hoc clustered networks," in *in Proc. of Military Communications and Information Systems Conference (MCC)*, Gdansk, Poland, Oct. 2012, pp. 1–8.

[127] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

Appendices

Appendix A

Proof of the equivalence between Definition 2.6 and the standard definition of the MNE

Let us define the mixed-strategy Nash equilibrium as follows [74]:

Definition A.1 (Nash equilibrium in Mixed strategy)**.** *A vector of probability distributions* $\boldsymbol{\pi} = (\pi_{1,a_1}, \ldots, \pi_{K,a_K})$ *is a Nash equilibrium in mixed strategy iff* $\forall\ k \in \mathcal{K},\ \forall\ a'_k \in \mathcal{A}_k$*:*

$$\bar{u}_k(\boldsymbol{\pi}_k,\ \boldsymbol{\pi}_{-k}) \geq \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k}). \tag{A.1}$$

Here $u_k(\boldsymbol{\pi}_k,\ \boldsymbol{\pi}_{-k})$ is the *expected* utility of player k when the strategy profile is $\boldsymbol{\pi}$, that is:

$$\bar{u}_k(\boldsymbol{\pi}_k,\ \boldsymbol{\pi}_{-k}) = \sum_{\boldsymbol{a}\in\mathcal{A}} u_k(\boldsymbol{a})\boldsymbol{\pi}(\boldsymbol{a}), \tag{A.2}$$

where $\boldsymbol{\pi}(\boldsymbol{a})$ is the probability that the players adopt the strategy profile $\boldsymbol{a}$, i.e, the probability that player 1 selects an action a_1 and player 2 selects an action a_2, etc. Since the players play in an independent way, then these events are independent, hence the probability $\boldsymbol{\pi}(\boldsymbol{a})$ equals to the product of the single probabilities:

$$\boldsymbol{\pi}(\boldsymbol{a}) = \prod_{j\in\mathcal{K}} \pi_{j,a_j}, \tag{A.3}$$

which gives us:

$$\bar{u}_k(\boldsymbol{\pi}_k,\ \boldsymbol{\pi}_{-k}) = \sum_{\boldsymbol{a}\in\mathcal{A}} u_k(\boldsymbol{a}) \prod_{j\in\mathcal{K}} \pi_{j,a_j}. \tag{A.4}$$

Using (A.4), it is possible to write (A.1) as

$$\sum_{\boldsymbol{a}\in\mathcal{A}} u_k(\boldsymbol{a}) \prod_{j\in\mathcal{K}} \pi_{j,a_j} \geq \sum_{\boldsymbol{a}_{-k}\in\mathcal{A}_{-k}} u_k(a'_{-k}, \boldsymbol{a}_{-k}) \prod_{j\in\mathcal{K}, j\neq k} \pi_{j,a_j}. \tag{A.5}$$

We claim that Definition A.1 and Definition 2.6 are equivalent.

Proof. In order to show this equivalence, we prove that a strategy profile is a MNE in the sense of one definition if and only if it is a MNE also in the sense of the other definition. We begin by showing that if π is a MNE in the sense of Definition 2.6, then it is also a MNE in the sense of Definition A.1. Since (2.5) is true $\forall\ a_k \in \mathcal{A}_k$, it is possible to sum both sides of the inequalities:

$$\sum_{a_k \in \mathcal{A}_k} \sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a_k, \boldsymbol{a_{-k}}) \prod_{j \in \mathcal{K}} \pi_{j,a_j} \geq \sum_{a_k \in \mathcal{A}_k} \sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \prod_{j \in \mathcal{K}} \pi_{j,a_j}. \tag{A.6}$$

On the left side, the two sums can be written as a single sum over the whole action set, while, on the right side, it is possible to exchange the order of the two sums, obtaining:

$$\sum_{\boldsymbol{a} \in \mathcal{A}} u_k(a_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j} \quad \geq \quad \sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \sum_{a_k \in \mathcal{A}_k} \prod_{j \in \mathcal{K}} \pi_{j,a_j} \tag{A.7}$$

$$\sum_{\boldsymbol{a} \in \mathcal{A}} u_k(a_k, \boldsymbol{a_{-k}}) \prod_{j=1}^{K} \pi_{j,a_j} \quad \geq \quad \sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \prod_{j \in \mathcal{K}, j \neq k} \pi_{j,a_j}, \tag{A.8}$$

where we used the well known formula:

$$\sum_{a_k \in \mathcal{A}_k} \prod_{j \in \mathcal{K}} \pi_{j,a_j} = \prod_{j \in \mathcal{K}, j \neq k} \pi_{j,a_j}. \tag{A.9}$$

Since (A.8) is equivalent to (A.5), this concludes the first part of the proof.

Now we show that if a strategy profile is a MNE in the sense of Definition A.1, then it is a MNE also in the sense of Definition 2.6. We begin by rewriting (2.5) with no loss of generality as:

$$\sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a_k, \boldsymbol{a_{-k}}) \pi_{k,a_k} \prod_{j \in \mathcal{K}, j \neq k} \pi_{j,a_j} \quad \geq \quad \sum_{\boldsymbol{a}_{-k} \in \mathcal{A}_{-k}} u_k(a'_k, \boldsymbol{a_{-k}}) \pi_{k,a_k} \prod_{j \in \mathcal{K}, j \neq k} \pi_{j,a_j} \tag{A.10}$$

$$\bar{u}_k(a_k, \boldsymbol{\pi}_{-k}) \pi_{k,a_k} \quad \geq \quad \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k}) \pi_{k,a_k}. \tag{A.11}$$

Notice that (A.11) is automatically true if $\pi_{k,a_k} = 0$, hence it is necessary to prove that (2.6) implies (A.11) only in the case in which $\pi_{k,a_k} > 0$.

Recall that at a MNE in the sense of Definition A.1, each player must be indifferent to any pure strategy to which it gives a positive probability[1], that is:

$$\bar{u}_k(a_k, \boldsymbol{\pi}_{-k}) \quad = \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k}) \quad \forall\ a_k, a'_k : \pi_{k,a_k} > 0, \pi_{k,a'_k} > 0 \tag{A.12}$$

$$\bar{u}_k(a_k, \boldsymbol{\pi}_{-k}) \quad > \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k}) \quad \forall\ a_k, a'_k : \pi_{k,a_k} > 0, \pi_{k,a'_k} = 0, \tag{A.13}$$

[1] This is sometimes known as the *indifference theorem*.

which means that $\forall\ a'_k$ and $\forall\ a_k : \pi_{k,a_k} > 0$

$$\bar{u}_k(a_k, \boldsymbol{\pi}_{-k}) \geq \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k}). \tag{A.14}$$

Multiplying both sides of the inequality by π_{k,a_k} we obtain that $\forall\ a'_k$ and $\forall\ a_k : \pi_{k,a_k} > 0$

$$\bar{u}_k(a_k, \boldsymbol{\pi}_{-k})\pi_{k,a_k} \geq \bar{u}_k(a'_k, \boldsymbol{\pi}_{-k})\pi_{k,a_k}. \tag{A.15}$$

Since (A.15) is equivalent to (A.11), thus it is equivalent to (2.6), this concludes the proof. □

Appendix B

Proof of theorem 4.1

Proof. Consider two arbitrary NE $\boldsymbol{a}^*$ and $\boldsymbol{a}^+ \in \mathcal{A}_{\text{NE}}$, such that $\sum_{\ell\in\mathcal{L}} \mathbb{1}_{\{\xi(\boldsymbol{a}^*)>\Gamma\}} = L^*$, $\sum_{\ell\in\mathcal{L}} \mathbb{1}_{\{\xi(\boldsymbol{a}^+)>\Gamma\}} = L^+$ with $L^* \geq L^+ + 1$. From Theorem 2.10, the stochastically stable points of the TE algorithm are the NE that maximize the social welfare W. Therefore, proving the theorem is equivalent to proving that $W(\boldsymbol{a}^*) > W(\boldsymbol{a}^+)$.

The social welfare associated with $\boldsymbol{a}^*$ using the utility in (3.7) is

$$\begin{aligned} W(\boldsymbol{a}^*) &= \sum_{k\in\mathcal{K}} u_k(\boldsymbol{a}^*) \\ &= \sum_{k\in\mathcal{K}} \frac{1}{1+\beta L_{\max}} \left(\varphi_k(\boldsymbol{a}^*) + \beta \sum_{\ell\in\mathcal{L}_k} \mathbb{1}_{\{\xi_i(\boldsymbol{a}^*)>\Gamma\}} \right) \\ &= \frac{1}{1+\beta L_{\max}} \left(\beta L^* + \sum_{k=1}^{K} \varphi_k(\boldsymbol{a}^*) \right). \end{aligned} \tag{B.1}$$

Since φ_k is a non-negative function, it holds that

$$W(\boldsymbol{a^*}) \geq \frac{\beta L^*}{1+\beta L_{\max}}. \tag{B.2}$$

Analogously, the social welfare associated with $\boldsymbol{a}^+$ is

$$\begin{aligned} W(\boldsymbol{a}^+) &= \sum_{k\in\mathcal{K}} u_k(\boldsymbol{a}^+) \\ &= \sum_{k\in\mathcal{K}} \frac{1}{1+\beta L_{\max}} \left(\varphi_k(\boldsymbol{a}^+) + \beta \sum_{\ell\in\mathcal{L}_k} \mathbb{1}_{\{\xi_i(\boldsymbol{a}^+)>\Gamma\}} \right) \\ &= \frac{1}{1+\beta L_{\max}} \left(\beta L^+ + \sum_{k=1}^{K} \varphi_k(\boldsymbol{a}^+) \right). \end{aligned} \tag{B.3}$$

By definition, $\forall\ \boldsymbol{a} \in \mathcal{A}^K$, and $\forall\ k \in \mathcal{K}$, $\varphi_k(\boldsymbol{a}) \leq 1$ and thus $W(\boldsymbol{a}^+) \leq \frac{\beta L^+ + K}{1+\beta L_{\max}}$. Then, using the assumption that $L^+ \leq L^* - 1$, it holds that

$$W(\boldsymbol{a}^+) \quad \leq \quad \frac{\beta L^* - \beta + K}{1 + \beta L_{\max}}.$$

Therefore, from the assumption that $\beta > K$ it is possible to write

$$\frac{\beta L^* - \beta + K}{1 + \beta L_{\max}} < \frac{\beta L^*}{1 + \beta L_{\max}}, \tag{B.4}$$

thus, following the chain of inequalities, it holds that $W(\boldsymbol{a}^+) < W(\boldsymbol{a}^*)$. This concludes the proof. □

Appendix C

Proof of theorem 4.2

Proof. From the assumptions of Theorem 4.2, the intersection between the set of NE $\mathcal{A}_{\text{NE}}$ and the set of solutions of (3.1) $\mathcal{A}^{\dagger}$ is non-empty, i.e., $\mathcal{A}_{\text{NE}} \cap \mathcal{A}^{\dagger} \neq \emptyset$. Let $\boldsymbol{a^*} \in \mathcal{A}_{\text{NE}} \cap \mathcal{A}^{\dagger}$ be an arbitrary element of the intersection and $L^* = \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{\xi(\boldsymbol{a}^*) > \Gamma\}}$ the number of links that satisfy their constraints. Since $\boldsymbol{a^*} \in \mathcal{A}^{\dagger}$ it results that $L^* = \max_{\boldsymbol{a} \in \mathcal{A}^K} \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{\xi(\boldsymbol{a}) > \Gamma\}}$, i.e., L^* is the maximum number of links that can simultaneously satisfy their constraints. From Theorem 2.10, the set of the stochastically stable action profiles is $\mathcal{A}_{\text{TE}} = \left\{ \boldsymbol{a}' \in \mathcal{A}^K : \boldsymbol{a}' \in \arg\max_{a \in \mathcal{A}_{\text{NE}}} W(\boldsymbol{a}) \right\}$. Hence, proving the theorem is equivalent to prove that $\mathcal{A}_{\text{TE}} \subseteq \left(\mathcal{A}_{\text{NE}} \cap \mathcal{A}^{\dagger} \right)$. From its definition $\mathcal{A}_{\text{TE}} \subseteq \mathcal{A}_{\text{NE}}$, thus it remains to prove that $\mathcal{A}_{\text{TE}} \subseteq \mathcal{A}^{\dagger}$.

Let $\mathcal{A}^{\star} \subseteq \mathcal{A}_{\text{NE}}$ be the set of NE such that $\forall\ \boldsymbol{a} \in \mathcal{A}^{\star}\ \sum_{\ell \in \mathcal{L}} \mathbb{1}_{\{\xi_\ell(\boldsymbol{a}) > \Gamma\}} = L^*$. Then, it results that $\forall\ \boldsymbol{a}^+ \in \mathcal{A}^K \setminus \mathcal{A}^{\star}$ it hold that $\sum_{\ell \in \mathcal{L}} \mathbb{1}_{\left\{\xi_\ell(\boldsymbol{a}^+) > \Gamma\right\}} < L^*$. Thus, from Theorem 4.1 and the assumption that $\beta > K$, it holds that $W(\boldsymbol{a}^+) < W(\boldsymbol{a})$, $\forall\ \boldsymbol{a}^+ \in \left(\mathcal{A}^K \setminus \mathcal{A}^{\star} \right)$ and $\forall\ \boldsymbol{a} \in \mathcal{A}^{\star}$. Therefore the set of stochastically stable points can be expressed as $\mathcal{A}_{\text{TE}} = \left\{ \boldsymbol{a}' \in \mathcal{A}^K : \boldsymbol{a}' \in \arg\max_{a \in \mathcal{A}^{\star}} W(\boldsymbol{a}) \right\}$. The social welfare of the action profiles on $\mathcal{A}^{\star}$ is:

$$W(\boldsymbol{a}) = \beta L^* + \sum_{k=1}^{K} \varphi_k(\boldsymbol{a}). \tag{C.1}$$

Therefore, $\arg\max_{\boldsymbol{a} \in \mathcal{A}^{\star}} W(\boldsymbol{a}) = \arg\max_{\boldsymbol{a} \in \mathcal{A}^{\star}} \sum_{k=1}^{K} \varphi_k(\boldsymbol{a})$. Thus, $\mathcal{A}_{\text{TE}}$ is the set of the action profiles that satisfy the constraints for L^* links and maximizes the $\sum_{k=1}^{K} \varphi_k(\boldsymbol{a})$, hence $\mathcal{A}_{\text{TE}} \subseteq \mathcal{A}^{\dagger}$. This concludes our proof. □

Appendix D

Markov Chain transition probabilities

D.1 Transition probability from an NE to a *discontent* state

The transition probability between an NE state and a state with one *discontent* player is denoted by $P(N, D)$. For the system to exit an NE, a player must pass from a *content* to a *discontent* state. This happens only in the following case: at time t player k experiments and during this experimentation k interferes with enough power to turn player l into watchful, at time $(t+1)$ player m experiments and during this experimentation m interferes turning l into *discontent*. The probability of at least one player experimenting in the system is given by: $P_\epsilon = 1-(1-\epsilon)^K$. By using the first two terms (reasonable since $\epsilon \ll 1$ implies $\epsilon^N \ll \epsilon^{(N-1)}$) of the binomial expansion $(1+(-\epsilon))^K = \sum_{k=0}^{K} \binom{K}{k}(-\epsilon)^k$ it holds that $P_\epsilon \simeq K\epsilon$. The probability that the player k disturbs another one, say l, is given by: (a) the probability of choosing an already occupied channel multiplied by (b) the probability of selecting a power level high enough. As a worst case scenario, assume that any power level greater than first quantization level is enough to create an intolerable level of interference. Thus, this probability is given by:

$$P_d = \underbrace{\frac{K-1}{C}}_{(a)} \underbrace{\frac{(Q-1)}{Q}}_{(b)}. \tag{D.1}$$

The probability that a player different from l experiments is $(K-1)\epsilon$, the probability of choosing the channel employed by l is $\frac{1}{C}$ and the probability of selecting a power level high enough is again given by (D.1). Therefore,

$$P(N, D) = K\epsilon\tfrac{(K-1)}{C}\tfrac{(Q-1)}{Q}(K-1)\epsilon\tfrac{1}{C}\tfrac{(Q-1)}{Q} \tag{D.2}$$

$$= \tfrac{K(K-1)^2\epsilon^2}{C^2}\left(\tfrac{Q-1}{Q}\right)^2. \tag{D.3}$$

D.2 Transition probability from *discontent* state to an NE

Here, we aim at evaluating $P(D,N)$, i.e., the transition probability between a state in which one player is *discontent* and a state in which all players are at the NE. Therefore, one player is performing a *noisy* search. Thus, the probability of immediately returning to an NE is given by: (a) the probability of selecting a free channel times (b) the probability of selecting enough power. Thus, we obtain:

$$P(D,N) = \underbrace{\frac{C-(K-1)}{C}}_{(a)} \underbrace{\frac{1}{Q}}_{(b)} . \tag{D.4}$$

D.3 Transition probability from a discontent state to a content state

The transition probability from a state with one *discontent* player to a state in which $K-k$ players are employing an individually optimal action is denoted by $P(D, C_{K-k})$. The *discontent* player selects a random action, then the probability of quitting the *discontent* state to a state in which only $(K-k)$ players are using one of their individually optimal actions depends on the acceptance function $F(u)$. Given (2.11) and for K large enough, the accepting probability can be approximated by $\epsilon^{F(u)} \approx 1$. When a player is *discontent*, it is possible for it to accept as a benchmark action the one that makes another player to change into a *discontent* mood. Then, the transition probability towards state C_{K-k} is given by the product of the probability of disturbing $(k-1)$ players that were at an NE before selecting a free channel or a channel used by a player that is not at an NE. The probability of colliding with $k-1$ players is given by

$$\frac{(K-1)}{C}\frac{(K-2)}{C}\frac{(K-3)}{C}...\frac{(K-k+1)}{C} = \frac{(K-1)!}{C^{k-1}\,(K-k)!}, \tag{D.5}$$

while the probability of selecting a channel free or used by a player not using an individually optimal action is $\frac{C-(K-k)}{C}$. Therefore, the product is:

$$P(D, C_{K-k}) = \frac{1}{C^k}\frac{(K-1)!}{(K-k)!}\,(C-K+k)\,. \tag{D.6}$$

D.4 Transition probability from C_{K-k} to C_{K-k+1}

The transition probability between a state in which $K-k$ players are using an individually optimal action and a state in which $K-k+1$ players are using an individually optimal action is denoted by $P\,(C_{K-k}, C_{K-k+1})$. Since no player is *discontent*, the transition happens through experimentation. To pass from a state in which $K-k$ players

are using an individually optimal action to another one in which $K-k+1$ are doing the same, the following sequence of events must happen: at least one of the $K-k$ players experiments; it selects one of the available individually optimal actions; and it accepts the action. Thus, the transition probability is

$$P\left(C_{K-k}, C_{K-k+1}\right) = \underbrace{(K-k)\epsilon}_{(a)} \underbrace{\frac{C-k}{CQ}}_{(b)} \underbrace{\epsilon^{G(\Delta u)}}_{(c)} \tag{D.7}$$

$$= (K-k)\frac{C-k}{CQ}\epsilon^{1+G(\Delta u)}. \tag{D.8}$$

Appendix E

Proof of Theorem 4.3

Proof. With a standard Markov chain analysis, starting from state C_0, the expected number of iterations before reaching for the first time the NE is given by: $\bar{T}_{NE} = \sum_{k=0}^{K-1} \frac{1}{P(C_{K-k}, C_{K-k+1})}$. Substituting, we obtain

$$\begin{aligned} \bar{T}_{NE} &= \frac{CQ}{\epsilon^{(1+G(\Delta u))}} \sum_{k=0}^{K-1} \frac{1}{(K-k)(C-k)} \\ &= \frac{CQ}{\epsilon^{(1+G(\Delta u))}(C-K)} \sum_{k=0}^{K-1} \left(\frac{1}{K-k} - \frac{1}{C-k} \right). \end{aligned} \tag{E.1}$$

Given (2.10) and the fact that $\epsilon \ll 1$, the following approximation holds $\epsilon^{(1+G(\Delta u))} \approx \epsilon$. For the sake of simplicity, in the following, the pre-multiplying constant factor is omitted and define $m = K - k$. Thus, equation (E.1) can be written as

$$\sum_{m=1}^{K} \left(\frac{1}{m} - \frac{1}{C-K+m} \right). \tag{E.2}$$

It is known that $\sum_{m=1}^{K} \frac{1}{m} < 1 + \int_1^K \frac{1}{x} dx$ thus:

$$\sum_{m=1}^{K} \frac{1}{m} \leq \log(K) + 1. \tag{E.3}$$

It is also known that the harmonic sum is such that

$$\sum_{m=1}^{K} \frac{1}{m} \geq \log(K) + \gamma. \tag{E.4}$$

Consider that $\forall n \geq 1$, with $K \in \mathbb{N}$ and $A \in \mathbb{N}$,

$$\int_n^{K+1} \frac{1}{A+x} dx < \sum_{m=n}^{K} \frac{1}{A+m} < \int_{n-1}^{K} \frac{1}{A+x} dx, \tag{E.5}$$

and thus, for the second addend it holds that:

$$\sum_{m=1}^{K} \frac{1}{C-K+m} \leq \log\left(\frac{C}{C-K}\right), \tag{E.6}$$

$$\sum_{m=1}^{K} \frac{1}{C-K+m} \geq \log\left(\frac{C+1}{C-K+1}\right). \tag{E.7}$$

By joining together equation (E.3) with (E.6) and (E.4) with (E.7), and by reinserting the omitted multiplicative factor, we obtain the result, and this concludes the proof. □

Appendix F

Proof of Theorem 4.4

Proof. The average fraction of time the system is at an NE can be expressed as $(1-\delta) = \frac{T_N}{T_{TOT}}$, where $\bar{T}_N$ is the expected time spent at an NE once it has been reached and by T_{TOT} the total time spent in all the states. Given the DTMC in Figure 4.2, this can be expressed as

$$T_{TOT} = \bar{T}_N + T_{BNE}, \tag{F.1}$$

where $\bar{T}_{BNE}$ denotes the expected time between the instant the system leaves an NE and the instant it reaches it again. The expected number of time steps needed to leave the NE once reached is

$$\begin{aligned}
\bar{T}_N = & \sum_{n=1}^{\infty} nP(NE,D)(1-P(NE,D))^{(n-1)} \\
= & -P(NE,D)\tfrac{d}{dP(NE,D)} \sum_{n=1}^{\infty}(1-P(NE,D))^n \\
= & -P(NE,D)\tfrac{d}{dP(NE,D)} \left(\tfrac{1}{P(NE,D)}\right) \\
= & \tfrac{1}{P(NE,D)}.
\end{aligned}$$

Here, the well known equality $\sum_{n=1}^{\infty} x^n = \frac{x}{1-x}$ has been used and $\sum_{n=1}^{\infty} nx^{(n-1)} = \frac{d}{dx}\sum_{n=1}^{\infty} x^n$. Thus, it follows that

$$(1-\delta) = \frac{1}{1+P(NE,D)T_{BNE}}. \tag{F.2}$$

To evaluate T_{BNE}, the process is as follows. The starting state on the Markov chain is the state D. From here, it is possible to go back to the NE state without quitting the discontent state. To do this, the expected number of time steps needed is $T_{(D,NE)} = \sum_{n=1}^{\infty} nP(D,N)P(D,D)^{(n-1)}$. These equalities imply the following

$$T_{(D,NE)} = \frac{P(D,NE)}{(1-P(D,D))^2}, \tag{F.3}$$

where $P(D, D)$ is easily obtained by imposing the sum of the probabilities to be equal to 1:

$$P(D, D) = 1 - \left(P(D, NE) + \sum_{k=1}^{K} P(D, C_{K-k}) \right). \tag{F.4}$$

On the other hand, it is possible to transit from the discontent state to a certain C_{K-k} state and the expected time steps needed to return to the NE starting from state C_{K-k} is denoted by $T_{CNE}(k)$. This quantity can be upper-bounded by using (E.5):

$$T_{CNE}(k) \leq \frac{CQ}{\epsilon^{1+G(\Delta u)}\,(C-K)} \left(\gamma + \log \left(\frac{K\,(C-k+1)}{C+1} \right) \right). \tag{F.5}$$

In the following, this upper bound is used as a close enough approximation of the true value. Moreover, given (2.10), and $\epsilon \ll 1$, it follows that $\epsilon^{1+G(\Delta u)} \approx \epsilon$. As consequence, the expected time T_{BNE} to return to an NE when the system deviates is given by:

$$T_{BNE} = T_{(D,NE)} + \sum_{k=1}^{K} P(D, C_{K-k}) T_{CNE}(k). \tag{F.6}$$

This concludes the proof. □