Learning Nash Equilibria in Distributed Channel Selection for Frequency-agile Radios

Irene Macaluso and **Luiz DaSilva** and **Linda Doyle**[1]

Abstract.

Wireless communication networks are evolving towards self-configuring, autonomous and distributed multiagent systems in which nodes are deployed randomly and have to adapt to the environment in which they operate. A cognitive network is a self-organising system that relies on the ability of its autonomous nodes to support communication in an adaptive and distributed manner. In this paper we address the distributed channel selection problem, which is a crucial component of many cognitive networks scenario, in the context of frequency-agile radios that are able to operate in multiple frequency bands simultaneously. We formulate the problem as an N-player stochastic game with incomplete information. We prove that by adopting a simple reinforcement scheme, namely learning automata, nodes will converge to a Nash equilibrium, under the assumption of symmetric interference between the players.

1 Introduction

Recent years have seen the evolution of traditional wireless networks towards self-configuring, autonomous and distributed multi-agent systems. This trend opens the doors to a new and exciting research field involving both the specialization of state-of-the-art general purpose multiagent algorithms and the development of new techniques to solve issues which are specific of the wireless communication domain.

A crucial factor in the evolution of wireless communication systems is the management of radio spectrum. In today's static communication systems, regulators decide both the use of a certain frequency band as well as the users that are allowed to transmit on that band (see Figure 1(a)). On the one hand this model, by rigidly planning the allocation of frequency bands, ensures that the interference between neighboring systems is limited. On the other hand, as the allocation of spectrum bands is implemented on a long term assignment basis, it is widely recognised that such an approach leads to an inefficient usage of the spectrum resources. A first consequence of this quasi-static paradigm is the presence of idle capacity within the system: numerous studies and measurement campaigns have shown that spectrum resources are often underutilized.

As spectrum becomes a more and more valuable resource, it is imperative to address the inherent inefficiency that characterizes the current spectrum management mechanism. Previous work focused on mechanisms that assign spectrum more dynamically over shorter time frames on an as-needed basis (see Figure 1(b)). The opening up of the TV white spaces [2] has made the concept of dynamic access to spectrum a reality. Networks using the TV white spaces can avail of any unsed spectrum in the TV bands as distinct from being assigned a static allowance. Frequency-agile radios make it possible to use whatever spectrum is available. In addition further advancements mean that radios can use non-contiguous as well as contiguous blocks of spectrum.

The use of non-contiguous blocks of spectrum for communications is increasingly of interest. LTE-Advanced [10] introduces the concept of carrier aggregation which facilitates the combining of different blocks of LTE spectrum to form larger transmission bandwidths. This will be crucial in supporting the ever-increasing data demands on mobile networks. In LTE-Advanced carrier aggregation can be performed using contiguous or non-contiguous blocks of spectrum and radio receiver architectures are being designed with this in mind. Though carrier aggregation, as currently envisaged, will involve the aggregation of static assignments of spectrum, it is not unrealistic to envisge this becoming more dynamic in the future.

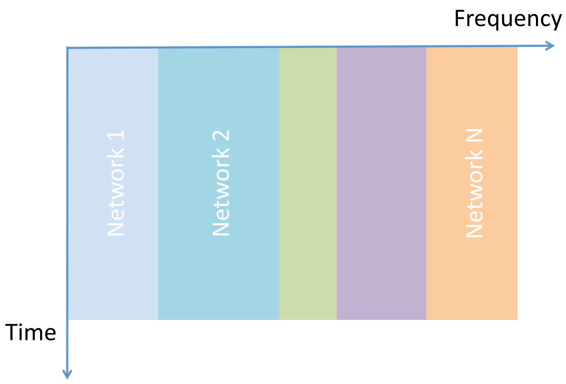
In this paper we look at scenarios in which a group of networks dynamically access non-contiguous blocks of spectrum and do so without the need for central coordination. We address the problem of distributed spectrum resources allocation in the context of frequency-agile radios that are able to operate in multiple frequency bands simultaneously. A number of frequency bands, herein named channels, have to be assigned to a number of wireless networks in a distributed manner so that the interference between adjacent systems is minimized. As each radio is able to operate simultaneously on multiple non-contiguous frequency bands, we extend the traditional distributed channel selection problem so that each network has to decide how many and which channels it should access. The problem of distributed channel selection has been addressed in the literature from a game theory point of view [7], from a distributed multi-agent learning perspective [1, 3] as well as a combined approach [8]. For special types of games, e.g. potential games, a proof of convergence of a reinforcement learning procedure has been provided [6]. We have recently applied learning automata to the problem of distributed channel selection for radios that have to decide whether it is advantageous to select an additional channel as opposed to keep using only its current transmission channel [4].

The problem we address in this paper is an extension of the traditional channel selection problem in that we allow each network to take advantage of of non-contiguous frequency bands. This problem can be formulated as an N-player stochastic game with incomplete information, i.e. the distribution of each player's payoff is unknown to other players. All the wireless networks have to reach, in an autonomous and distributed manner, a stable allocation of the available channels, corresponding to a Nash equilibrium (NE). We model each

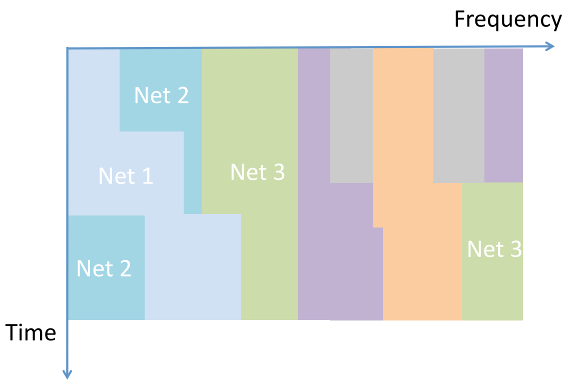
[1] CTVR Telecommunications Research Centre, Trinity College Dublin, Ireland

player as a learning automaton [5], which is a reinforcement learning scheme where each agent is a policy iterator. We prove that by adopting this simple reinforcement scheme players will converge to a Nash equilibrium, under the assumption of symmetric interference between the players.

The remainder of the paper is organized as follows. Section 2 details the system model and some of the key properties of learning automata. Section 3 provides the proof of convergence to a NE of the proposed learning procedure. Section 4 presents the numerical results. We summarize our conclusions in section 5.



(a) Fixed band planning.



(b) Dynamic channel allocation.

Figure 1. (a) A central entity divides the spectrum into frequency channels which are then assigned to various technology/users. (b) Spectrum resource assignment changes in time according to the requirements of the various wireless networks.

2 Problem Description

Let us assume that N wireless networks are operating in the same geographical area where M frequency channels are available. Two or more networks can share the same channel only if the interference they cause to each other is below a certain threshold; otherwise they have to choose different channels. The goal of a distributed channel allocation algorithm is to compute a stable channel assignment, either through negotiation or blind interaction between the networks, so as to minimize interference between the systems.

In this paper we adopt the blind interaction approach. The players indirectly interact by causing interference to each other. No explicit communication is allowed. This tight restriction is a desideratum when one is designing a communication system in which cooperation between the different players cannot be assumed. Moreover it simplifies the system design: in the case of cooperation between networks a common control channel to exchange information has to be established and maintained; also different networks need to agree on a common protocol. Thus these requirements might hinder the deployment of heterogenous wireless networks.

We model each player i as a learning automaton: each player i keeps a probability distribution $\mathbf{p}_i(k)$ over the set of actions and, at each stage k, it chooses an action according to $\mathbf{p}_i(k)$. Then each player updates its action probabilities based on the response $r_i(k)$ from the environment, ignoring the presence of the other agents in the same environment.

In this paper we adopt the Linear Reward-Inaction scheme (L_{R-I}). The automaton increments the probability of the action that resulted in a favorable response from the environment, while accordingly decreasing the probabilities of all the other actions. In case of an unfavorable environment response, the action probabilities are left unaltered. The L_{R-I} updating rule is:

$$\mathbf{p}_i(k+1) = \mathbf{p}_i(k) + br_i(k)(\mathbf{e}_i - \mathbf{p}_i(k)) \tag{1}$$

where $b \in (0,1)$ is the reward parameter and $\mathbf{e}_i$ is the unit vector, where the i^{th} element is unity.

Each network i wants to get assigned a number of channels $c_i \leq M_i$, where $M_i \leq M \; \forall i$. Thus each agent has to decide how many and which channels it should access. By exploiting the development of frequency-agile radios, which are capable of transmitting on non-contiguous frequency bands, the cardinality of each agent's action space is:

$$|A_i| = \sum_{j=1}^{M_i} \binom{M}{j}. \tag{2}$$

In other words, each agent can decide to transmit on any of the possible combinations of the desired number of channels.

Let r_i be the payoff of player i, modeled as a random variable, with $r_i \in [0,1]$. Let us define the utility function of player i:

$$d^i(a_1, a_2, ..., a_N) = E[r_i | a_1, a_2, ..., a_N] \tag{3}$$

where a_j is the action chosen by player j. If we denote by p_{ij} the probability that player i chooses action j, a strategy for player i is defined as $\mathbf{p}_i = [p_{i1}, p_{i2}, ..., p_{i|A_i|}]$. As in [9], we define the functions f_{is} as the expected payoff of player i, given that player i selects action s and player l adopts strategy $\mathbf{p}_l$:

$$f_{is}(Q) = \sum_{\substack{j_1,\ldots,j_{i-1},\\ j_{i+1},\ldots,j_N}} d^i(j_1, \ldots, j_{i-1}, s, j_{i+1}, \ldots, j_N) \prod_{l \neq i} p_{l j_l} \tag{4}$$

where $Q = (\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N)$ and j_l is the action selected by player l with probability $p_{l j_l}$.

If the learning factor b in (1) is small the following statements hold [9]:

1. If the learning algorithm converges, it always converges to a NE.
2. For any N-player game, all strict Nash equilibria in pure strategies are asymptotically stable.
3. If there exists a bounded differentiable function $F : R^{M_1+M2+\ldots+M_N} \rightarrow R$, such that for some constant $v > 0$

$$\frac{\partial F}{\partial p_{is}}(Q) = v f_{is}(Q), \forall i, s \text{ and } \forall Q \in I^{M_1+M2+\ldots+M_N} \tag{5}$$

where $I = [0,1]$, then the learning algorithm always converges to an NE, for any initial condition.

3 Learning Nash Equilibria

We model the problem of multiple channel selection as an N-player game. Each player has to decide how many and which channel it should transmit on while minimizing the interference caused to adjacent systems. Thus a single action a_i denotes the set of channels selected by player i.

In our model the players' payoff is a function of each node's SINR. The SINR for player i transmitting on channel c_i is defined as:

$$\gamma_{ic_i} = \gamma_i(a_1, ..., c_i, ..., a_N) = \frac{|h_i(c_i)|^2 P}{\sigma_i^2(c_i) + \sum\limits_{j|c_i \in a_j} |h_{ji}(c_i)|^2 P} \quad (6)$$

where a_j is the action selected by player j, P is the transmit power, $|h_i(c_i)|$ is the channel gain between transmitter i and receiver i in channel c_i, $|h_{ji}(c_i)|$ is the channel gain between transmitter j and receiver i in channel c_i, and $\sigma_i^2(c_i)$ is the noise power at the receiver i in channel c_i. In our study we assume that the channel gains are time-invariant realizations of a circular symmetric complex normal distribution with zero mean and unit variance. The noise vector is also modeled as a circular symmetric complex Gaussian random variable.

We define the payoff of player i as:

$$r_i(a_1, ..., a_i, ..., a_N) = \frac{1}{M_i} \sum_{c_i \in a_i} \left(1 - \frac{\bar{\gamma}}{\gamma_{ic_i}}\right) + \left(1 - \frac{n_{a_i}}{M_i}\right) \quad (7)$$

where $\bar{\gamma}$ is the minimum SINR that receiver i can support, n_{a_i} is the number of channels that node i transmits on when selecting action a_i (i.e. $n_{a_i} = |a_i|$) and M_i is the maximum number of channels that node i is interested in.

Theorem 3.1. *A pure strategy NE $(a_1^*, .., a_i^*, ..., a_N^*)$, according to the payoff given in (7), corresponds to the players choosing the maximum number of channels characterized by an SINR which is greater than $\bar{\gamma}$.*

Proof. If action a_i^* includes a channel k for which the receiver i experiences an SINR $\gamma_{ik} \le \bar{\gamma}$, then the payoff of player i is:

$$\begin{aligned} r_i(a_1^*, .., a_i^*, ..., a_N^*) &= \frac{1}{M_i} \sum_{c_i \in a_i} \left(1 - \frac{\bar{\gamma}}{\gamma_{ic_i}}\right) + \left(1 - \frac{n_{a_i}}{M_i}\right) \\ &\le \frac{1}{M_i} \sum_{c_i \in \bar{a_i}} \left(1 - \frac{\bar{\gamma}}{\gamma_{ic_i}}\right) + \left(1 - \frac{n_{a_i} - 1}{M_i}\right) \\ &= r_i(a_1^*, .., \bar{a_i}, ..., a_N^*) \end{aligned} \quad (8)$$

where $\bar{a_i} = a_i \setminus \{k\}$.

If there exists a channel $k \notin a_i^*$ such that $\gamma_{ik} > \bar{\gamma}$, then:

$$\begin{aligned} r_i(a_1^*, .., \hat{a_i}, ..., a_N^*) &= \frac{1}{M_i} \sum_{c_i \in \hat{a_i}} \left(1 - \frac{\bar{\gamma}}{\gamma_{ic_i}}\right) + \left(1 - \frac{n_{\hat{a_i}}}{M_i}\right) \\ &> r_i(a_1^*, .., a_i^*, ..., a_N^*) \end{aligned} \quad (9)$$

where $\hat{a_i} = a_i \cup \{k\}$ and $n_{\hat{a_i}} = n_{a_i} + 1$.

□

To simplify the notation, we denote by $g_{ji}^{(s)}$ and $g_i^{(s)}$ the square of the expected value of the channel gain between player j and player i corresponding to channel s and the square of the expected value of the channel gain between the i^{th} transmitter-receiver pair in channel s. Let us define the *interference functions* as:

$$r_{ij}(c_i, c_j) = \begin{cases} \dfrac{g_{ji}^{(c_i)}}{g_i^{(c_i)}} & \text{if } c_i = c_j \text{ and } j \in N_i \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where N_i denotes the set of players that can interfere with player i. The interference functions are said to be symmetric if $r_{ij}(c_i, c_j) = r_{ji}(c_j, c_i), \forall i, j, a_i, a_j$.

Theorem 3.2. *The learning algorithm specified by (1) with payoff function given in (7) always converges to a NE if the interference functions are symmetric and the parameter b in (1) is small.*

Proof. Accordingly to the payoff function in (7), (4) can be rewritten as:

$$\begin{aligned} f_{is}(Q) &= \frac{1}{M_i} \sum_{c \in s} \left(1 - \frac{\bar{\gamma}}{g_i^{(c)} P} \left(\sigma^2(c) + \sum_{j \in N_i} g_{ji}^{(c)} P \sum_{a_j | c \in a_j} p_{ja_j}\right)\right) \\ &+ \left(1 - \frac{n_s}{M_i}\right) \end{aligned} \quad (11)$$

Let us define the function:

$$\begin{aligned} F(Q) &= \frac{2}{M_i} \sum_i \sum_k n_k p_{ik} - \frac{2\bar{\gamma}}{M_i P} \sum_i \sum_k \sum_{c_k \in k} \frac{\sigma^2(c_k)}{g_i^{(c_k)}} p_{ik} \\ &- \frac{\bar{\gamma}}{M_i} \sum_i \sum_j \sum_k \sum_t \sum_{c_k \in k} \sum_{c_t \in t} r_{ij}(c_k, c_t) p_{ik} p_{jt} \\ &+ 2 \sum_i \sum_k (1 - \frac{n_k}{M_i}) p_{ik} \end{aligned} \quad (12)$$

From (11) and (12), by exploiting the symmetry of the interference functions, we get:

$$\frac{\partial F}{\partial p_{is}}(Q) = \frac{2n_s}{M_i} - \frac{2\bar{\gamma}}{M_i} \sum_{c \in s} \left(\frac{\sigma^2(c)}{g_i^{(c)} P} + \frac{\sum\limits_{j \in N_i} g_{ji}^{(c)} \sum\limits_{a_j | c \in a_j} p_{ja_j}}{g_i^{(c)}} \right) + 2(1 - \frac{n_s}{M_i})$$

Therefore, according to the results reported in the previous section (theorems 3.2 and 3.3 in [9]), the learning algorithm always converges to a NE. □

4 Simulation results

In this section we discuss the effect the number of networks N, number of channels M, and noise power σ^2 has on the convergence time of the proposed learning algorithm. The payoff defined in (7) has a direct implementation and it allows a fully distributed solution to the problem of channel selection. The receiver estimates the SINR on the set of channels selected by its transmitter and sends back this information. The transmitter then updates its policy accordingly. If the transmitter does not receive an ACK from its receiver(s), an unfavorable response is assumed. For each scenario we run 10^5 independent simulations. We assumed that two or more networks cannot operate on the same channels.

Figure 2 shows the average number of iterations required to converge to a NE with respect to the cardinality of each player action space. We assumed that each network tries to access at most $M_i = 2$ channels. Thus the cardinality of the action space is $A_i = M + \binom{M}{2}$. We can observe that the convergence time increases with the number of players N. The impact of the cardinality of the action space A_i on the convergence time is an interesting aspect of these results. As expected, until a certain point the convergence time of the algorithm increases with A_i. Then, increasing the number of available actions does not affect the convergence time.

In Figure 3 we analyzed the combined effect of the number of available channels M and the noise power on the average number of iterations to converge to a NE. In particular, we assumed that $N = 4$ networks operate on the same geographical area and cannot utilize the same channels. Accordingly to what we discussed above, the impact of increasing the number of channels becomes less significant after a certain point, independently on the noise power. For each value of A_i, the convergence time linearly increases with the noise power.

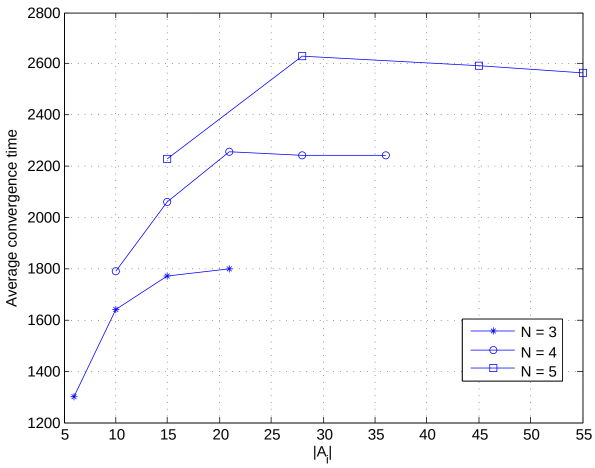


Figure 2. The average number of iterations to converge to a NE versus cardinality of each player action space ($\sigma^2 = 0.01$). For each scenario, i.e. number of players and number of available channels, we run 10^5 independent simulations ($b = 0.05$).

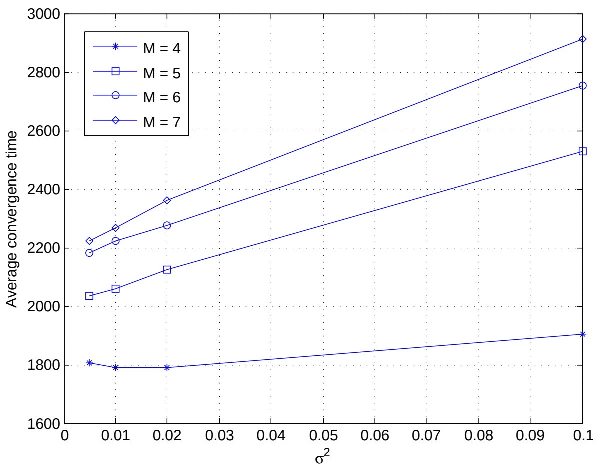


Figure 3. The average number of iterations to converge to a NE versus noise power at the receiver with $N = 4$ players. For each scenario, i.e. noise power and number of available channels M, we run 10^5 independent simulations ($b = 0.05$).

5 Conclusions

In this paper, we studied the problem of distributed channel allocation by exploiting a radio's ability to operate in multiple channels simultaneously. In particular, we extended the traditional channel selection problem to allow a radio to decide how many and which channels it should use, while minimizing the interference suffered from nearby systems. By modeling each player as a learning automaton and using the Linear Reward Inaction scheme, we formally proved convergence to a pure Nash equilibrium, under the assumption of symmetric interference between the players.

REFERENCES

[1] A. Anandkumar, N. Michael, A.K. Tang, and A. Swami, 'Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret', *Selected Areas in Communications, IEEE Journal on*, **29**(4), 731–745, (2011).

[2] Federal Communications Commission. Second report and order and memorandum opinion and order in the matter of unlicensed operation in the tv broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band, 14th November 2008.

[3] H. Li, 'Multiagent Q-learning for Aloha-like Spectrum Access in Cognitive Radio Systems', *EURASIP Journal on Wireless Communications and Networking*, **2010**, 56, (2010).

[4] I. Macaluso and L.A. DaSilva, 'Learning Automata and Distributed Channel Selection in Cognitive Networks', (under review, 2012).

[5] K.S. Narendra and M.A.L. Thathachar, *Learning automata: an introduction*, Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.

[6] S.M. Perlaza, H. Tembine, and S. Lasaulce, 'How can Ignorant but Patient Cognitive Terminals Learn their Strategy and Utility?', in *Signal Processing Advances in Wireless Communications (SPAWC), 2010 IEEE Eleventh International Workshop on*, pp. 1–5. IEEE, (2010).

[7] L. Rose, S.M. Perlaza, and M. Debbah, 'On the Nash Equilibria in Decentralized Parallel Interference Channels', in *IEEE Worshop Game Theory and Resource Allocation for 4G*, (2011).

[8] L. Rose, S.M. Perlaza, S. Lasaulce, and M. Debbah, 'Learning Equilibria with Partial Information in Decentralized Wireless Networks', *IEEE communications Magazine*, **49**(8), (2011).

[9] PS Sastry, VV Phansalkar, and MAL Thathachar, 'Decentralized Learning of Nash Equilibria in Multi-Person Stochastic Games with Incomplete Information', *Systems, Man and Cybernetics, IEEE Transactions on*, **24**(5), 769–777, (1994).

[10] Guangxiang Yuan, Xiang Zhang, Wenbo Wang, and Yang Yang, 'Carrier aggregation for lte-advanced mobile communication systems', **48**, 88–93, (2010).