Achieving Pareto Optimal Equilibria in Energy Efficient Clustered Ad Hoc Networks

Luca Rose, Samir Medina Perlaza, Christophe J. Le Martret, Mérouane Debbah

Achieving Pareto Optimal Equilibria in Energy Efficient Clustered Ad Hoc Networks

Luca Rose[1,3], Samir M. Perlaza[2], Christophe J. Le Martret[3], and Mérouane Debbah[1]
[1] Alcatel-Lucent Chair in Flexible Radio, SUPÉLEC, Gif-sur-Yvette, France, email: {luca.rose,merouane.debbah}@supelec.fr
[2] Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, email: perlaza@princeton.edu
[3] Thales Communications & Security, Gennevilliers , France, email: christophe.le_martret@thalesgroup.com

***Abstract*—In this paper, a decentralized iterative algorithm, namely the optimal dynamic learning (ODL) algorithm, is analysed. The ability of this algorithm of achieving a Pareto optimal working point exploiting only a minimal amount of information is shown. The algorithm performance is analysed in a clustered *ad hoc* network, where radio devices are assumed to operate above a minimal signal to interference plus noise ratio (SINR) threshold while minimizing the global power consumption. Sufficient analytical conditions for ODL to converge to the desired working point are provided, moreover through numerical simulations the ability of the algorithm to configure an interference limited network is shown. The performances of ODL and of a Nash equilibrium reaching algorithm are numerically compared, and their performance as a function of available resources is studied. The gain of ODL is shown to be larger when the amount of available radio resources is scarce.**
***Keywords*: Learning, power control, trial and error, Nash equilibrium, Pareto optimality, *ad hoc* network, channel selection, spectrum sharing.**

I. Introduction

In this paper, we consider a clustered *ad hoc* network where clusters share a given amount of logical channels, each cluster being allocated one channel at a time. Here, the clusters select their radio settings (which we assume to consist of a communication channel and a transmission power level) in a fully distributed manner without cooperation nor coordination. In this way, there is no need for inter-cluster information exchange which requires extra resources. Moreover, this allows the clusters to accommodate unforeseen situations where the spectrum may not be fully mastered. From a practical point of view, we consider the communication to be locally centralized through the election of a cluster head (CH) that, for the nodes within the cluster, takes care of the resource allocation, including transmit channel and power level. We assume the absence of any infrastructure which coordinates the CHs, thus interference may occur in between nodes belonging to different clusters which eventually select the same channel. As a consequence, the transmission becomes more unstable and problematic when the amount of available radio resources (e.g., frequency channels) is much smaller than the amount of CHs and nodes.

We propose an algorithm, namely optimal dynamic learning (ODL), which is able to select a Pareto optimal network working point. Along with the channel assignment problem, we also want to minimize the total transmit power over the network, to minimize both the interference level on the field and the battery drain of the nodes. This transmit power minimization is constrained by imposing a certain quality of service (QoS) per link within all clusters. Each CH is thus assumed to choose independently a resource configuration to meet link qualities expressed in terms of signal to interference plus noise ratio (SINR). The closest works to ours are [1]–[3]. In [1], an algorithm, namely trial and error (TE), was introduced and studied for achieving in a decentralized way an efficient equilibrium point; in [2], algorithms for a centralized power control based on variational inequality theory are developed; in [3], iterative water-filling is used to guarantee a certain achievable rate and the authors provide, in a low interference regime, a sufficient condition for the convergence. It is worth noting that in the works in [2] and [3] the power levels are taken from a compact set. Conversely, in our work, we consider finite power levels, since, in realistic networks, it must be expressed in a finite amount of bits. In addition, all the previous works present algorithms which aim at achieving a Nash equilibrium (NE). The reason to do so is that it is considered an intrinsically stable working point. The price to pay, however, is the eventual inefficiency of the equilibrium, as shown in [4]–[6]. In our approach, the network is not driven to reach a NE, but to reach a Pareto optimal working point. Our contributions in this paper are: (i) we describe a completely decentralized algorithm, namely an optimal dynamic learning (ODL) algorithm, able to keep the SINR level above a certain threshold a high proportion of the time; (ii) we prove that through our utility function, only *one bit* feedback per receiver and on local information are needed to reach a Pareto optimal solution (iii) we compare through numerical simulations the performance of ODL with Trial and error (TE), an algorithm that implements a Nash equilibrium solution with high probability. The paper is organized as follows. In section II we present and detail the two system models considered in this work; in section III, we introduce a game theoretical formulation for the network; in IV we briefly describe ODL and we state some theoretical results; in V we present and comment the results of our experiments; finally in VI we draw our conclusions.

II. System model

In this work, we consider two system models. The first is a simple mathematical abstraction often employed in the literature and known as interference channel (IC), used to illustrate some properties of our algorithm. The second scenario is a more realistic model of a dense network, which serves us to show the ability of the proposed algorithm to organize the transmission.

Let us define as $\mathcal{K} \triangleq \{1,\dots,K\}$ the set of clusters, $\mathcal{I} = \{1,2,\dots,I\}$ the set of transmitter-receiver pairs and $\mathcal{N}_k = \{1,2,\dots N_k\}$ the set of pairs belonging to cluster $k \in \mathcal{K}$

such that $\cup_{k \in \mathcal{K}} \mathcal{N}_k = \mathcal{I}$. Let us denote by $\mathcal{C} \triangleq \{1, 2, \ldots, C\}$ the set of channels that can be chosen by the CHs, let be $c_k \in \mathcal{C}$ the channel chosen by CH k. Each of these channels is then partitioned into B sub-channels assigned by the CH to the transmitters to communicate, we indicate by $\mathcal{B} = \{1, 2, \ldots, BK\}$ the sub-channels set and by $b_i \in \mathcal{B}$ the sub-channel used by the pair $i \in \mathcal{I}$ to communicate. Moreover, each transmitter i belonging to cluster k uses a transmission power p_k which is set by the CH. We denote by $\mathcal{P} \triangleq \{0, ..., P_{MAX}\}$ the set of available power levels such that $\forall k \in \mathcal{K}$ $p_k \in \mathcal{P}$ and $|\mathcal{P}| = Q$, i.e., Q is the number of quantization levels.

In the following, we denote by $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$ the network power allocation vector, by $\boldsymbol{c} = (c_1, c_2, \ldots, c_K)$ the spectrum occupation vector, by $\boldsymbol{a} = (a_1, a_2, ..., a_K)$ the network configuration vector, or *action profile*, where $a_k = (p_k, c_k)$ and $a_k \in \mathcal{A}_k = \mathcal{P} \times \mathcal{C}$. At every time instant, each transmitter attempts to communicate with its corresponding receiver which in turn evaluates the received signal's SINR. If it is greater than the minimum SINR threshold Γ, then the transmission was successful and the receiver transmits a positive ACK to the CH, otherwise, it transmits a NACK. Note that this notification mechanism requires only 1 bit feedback per receiver.

Here, the CHs have no information on the behaviour of the other clusters, thus they completely ignore other clusters' codebook or transmission scheme, and receivers treat interference as Gaussian noise. The multiple access interference (MAI) suffered from the receiver of the i-th pair belonging to cluster k is then:

$$MAI_i(\boldsymbol{a}) = \sum_{\ell \in \mathcal{K} \setminus k} \sum_{m=1}^{N_\ell} p_\ell G_{(m,i)}^{(b_i)} \mathbb{1}_{\{b_i = b_m\}}, \tag{1}$$

where $G_{(\ell,i)}^{(b_i)}$ represents the channel power gain between the transmitter of the l-th pair and the receiver of the i-th pair, over sub-band b_i and $\mathbb{1}_{\{\}}$ represents the indicator function. We assume the channel to be block-fading, i.e., it remains invariant during the whole transmission time. Given (1), the SINR achieved from the pair i belonging to cluster k is

$$SINR_i(\boldsymbol{a}) = \frac{p_k G_{(i,i)}^{(b_i)}}{\sigma^2 + MAI_i}, \tag{2}$$

where $G_{i,i}^{(b_i)}$ represents the channel power gain between the transmitter and the receiver of the pair i, employing the sub-band b_i and σ^2 is the power of the thermal noise, assumed to be constant over the whole spectrum.

Our objective is the satisfaction of the SINR constraints for the largest possible set of pairs by using the lowest global energy consumption. Formally, we want the network configuration vector $\boldsymbol{a}^*$ to be a solution of the following optimization problem

$$\begin{cases} \min\limits_{\boldsymbol{p} \in \mathcal{P}^K} \sum\limits_{k=1}^{K} p_k \\ \text{s.t. } SINR_i(\boldsymbol{a}) > \Gamma \quad \forall k \in \mathcal{K}^* \end{cases}, \tag{3}$$

where we denote by $\mathcal{K}^* \subseteq \mathcal{K}$ the largest subset of links able to simultaneously achieve a sufficient SINR level. Generally, to achieve this goal a central controller knowing all the network's parameters is required.

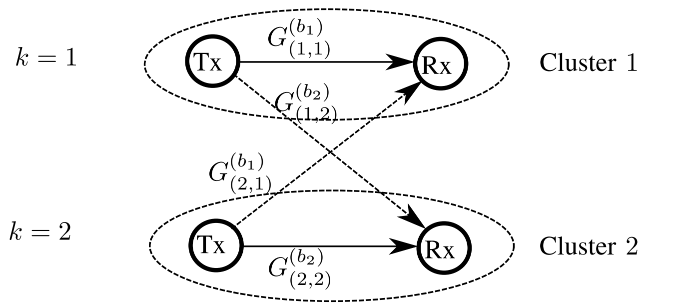

Fig. 1. System model for the interference channel. Here the two clusters share $C = 2$ available channels.

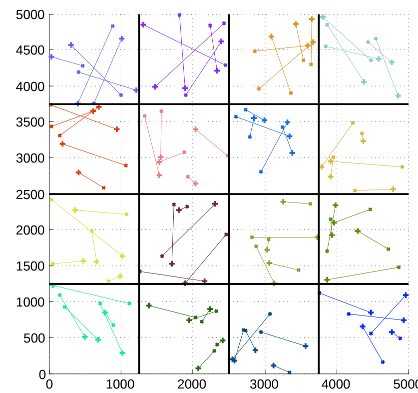

Fig. 2. Scenario of a dense system. Here, a square field with a 5 Km side, is populated with $K = 16$ square clusters each equipped with $N_k = 4$ transmitter-receiver pairs. Clusters share $C = 4$ channels choosing between $Q = 8$ power levels. The channel power attenuation is evaluated following a 2-ray model.

A. Small Networks

Here, we describe the system represented in Fig. 1. It is composed of $K = 2$ clusters, each populated with $N_k = 1$ transmitter-receiver pair, sharing the same spectrum divided into $C = 2$ orthogonal channels. Each transmitting device has a maximum output power of $P_{MAX} = 20$ W, linearly quantized into $Q = 3$ levels and tries to achieve a SINR level of $\Gamma = 10$ dB. Here, we consider a particular channel gain matrix G defined as:

$$G^{(1)} = \begin{bmatrix} 1 & 1 \\ 0.1 & 1 \end{bmatrix}, \quad G^{(2)} = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}. \tag{4}$$

We consider a normalized noise power, i.e., $\sigma^2 = 1$ W.

B. Large Networks

We describe the scenario represented in Fig. 2. The system is composed of $K = 16$ square clusters, divided on a square field of 5 Km side. Each cluster is populated by N_k transmitter-receiver pairs randomly positioned inside the clusters. For the sake of simplicity, we take N_k to be independent from k, i.e., $N_k = N$. Each transmitting device can transmit a maximum of $P_{MAX} = 50$ W logarithmically quantized into $Q = 8$ levels.

Here, each CH must choose between $C = 4$ available channels, each of which is assumed to have a passing bandwidth of $B = 1.25$ MHz, around a central frequency $f_0 = 400$ MHz. Channel power gain is evaluated through the well known 2-ray model [7]:

$$G_{(t,r)}^{b_i} = \frac{h_t^2 h_r^2 g_t g_r}{d(t,r)^4}, \quad \forall i, t, r \in \mathcal{I} \tag{5}$$

where $h_t = 1.5$ m and $h_r = 1.5$ m are, respectively, the transmitter and receiver antennas heights, $g_t = 1.5$ dB and $g_r = 1.5$ dB are, respectively, the transmitter and receiver antennas gain, and $d(t,r)$ is the distance between node t and node r.

III. Game Formulation

In this section, we briefly model the scenarios presented in Sec II under a normal-form formulation. Thank to this formulation, we are able to design a utility function that permits us to predict the most probable state of the network.

A game in normal-form is described by the triplet:

$$\mathcal{G} = \left(\mathcal{K}, \mathcal{A}, \{u_k\}_{k\in\mathcal{K}}\right), \tag{6}$$

where $\mathcal{K}$ represents the set of players, i.e. the CHs, $\mathcal{A}$ is the joint set of actions, that is, $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_K$ and we introduce the utility function $u_k : \mathcal{A} \to \mathbb{R}$ defined by:

$$u_k(\boldsymbol{a}) = \frac{1}{1+\beta N_k}\left(1 - \frac{p_k}{P_{MAX}} + \beta \sum_{i\in\mathcal{N}_k} \mathbb{1}_{\{SINR_i(\boldsymbol{a})>\Gamma\}}\right), \tag{7}$$

where β is a design parameter discussed in Sec IV. This function is an extension of the utility function introduced in [1] to a case with multiple pairs per cluster. To measure the global performance of an action profile we use the *social welfare*, defined by the sum of all individual utilities: $W(\boldsymbol{a}) = \sum_{k=1}^{K} u_k(\boldsymbol{a})$. Notice that, every action profile $\boldsymbol{a}^*$ such that $\boldsymbol{a}^* \in \arg\max_{\boldsymbol{a}\in\mathcal{A}} W(a)$ is also Pareto optimal.

In order to evaluate (7), each player, i.e. each CH, only requires intra-cluster information. In detail, the power level p_k is set by the CH and, each receiver i feeds back $\mathbb{1}_{\{SINR_i(\boldsymbol{a})>\Gamma\}}$ to the CH. We underline that this feedback mechanism only requires one bit per transmission.

In the following, we show that by using the utility function defined above, the action profile which maximize the social welfare of the game $\mathcal{G}$ solves the problem stated in (3), regardless of the underlying network topology.

The algorithm which we propose, is based on the concept of *interdependent* game, i.e. a game where the utility of any group of players depends on the action selected by at least one other player. More formally we can define an interdependent game as follows.

Definition 1: (Interdependent game). Let $\mathcal{G}$ be such that for every non-empty subset $\mathcal{K}^+ \subset K$ and every action profile $\boldsymbol{a} = (\boldsymbol{a}_{\mathcal{K}^+}, \boldsymbol{a}_{-\mathcal{K}^+})$[1] such that $\boldsymbol{a}_{\mathcal{K}^+}$ is the action profile of all players in $\mathcal{K}^+$, it holds that:

$$\exists i \notin \mathcal{K}^+, \exists \boldsymbol{a}'_{\mathcal{K}^+} \neq \boldsymbol{a}_{\mathcal{K}^+} : u_i(\boldsymbol{a}'_{\mathcal{K}^+}, \boldsymbol{a}_{-\mathcal{K}^+}) \neq u_i(\boldsymbol{a}_{\mathcal{K}^+}, \boldsymbol{a}_{-\mathcal{K}^+}), \tag{8}$$

then, it is interdependent.

In the following, we assume the game $\mathcal{G}$ to be interdependent. From a physical stand point, this means that in the network there is no cluster, or group of clusters, able to create interference without suffering any. This is a reasonable assumption in a wireless context if channel reciprocity holds. Moreover, in the system model introduced in section II-B this is guaranteed from the assumption of dealing with a dense network.

IV. Algorithm Description

In this section, we briefly describe the distributed learning algorithm introduced in [8] and we analyse its general properties.

A. Algortihm Description

In ODL, every player k implements a state machine, where a state $Z_k(n) = (m_k(n), a_k(n), u_k(n))$ is defined by a triplet composed by a *mood* $m_k(n)$, a *benchmark* utility $\bar{u}_k(n)$ and a *benchmark* action $\bar{a}_k(n)$. 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 n player k is content, it chooses action $a_k(n)$ following the probability distribution

$$\pi_{k,a_k} = \begin{cases} \frac{\epsilon^{K+1}}{|\mathcal{A}_k|-1} & \text{if} \quad \bar{a}_k \neq a_k \\ 1-\epsilon^{K+1} & \text{if} \quad \bar{a}_k = a_k. \end{cases}, \tag{9}$$

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

- **Discontent**

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

B. Algorithm properties

The algorithm previously described shows some useful properties shown in [8]; for the sake of simplicity, we rewrite the main result within with our notation.

Theorem 1: Let 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 [9], is at the base of the algorithm. Basically, a stochastically stable action profile is an action profile that, once it is reached by the

[1]Here, $a_{-\mathcal{K}^+}$ refers to the action profile of all the players that are not in $\mathcal{K}^+$

TABLE I
TABLE SUMMARIZING THE PERFORMANCE COMPARISON BETWEEN ODL AND TE ON A SMALL NETWORK.

	Average Welfare	Average Satisfaction
ODL	0.39	0.37
TE	0.21	0.11

algorithm, there is a small probability of leaving it. Note that, compared with other results in the literature, for instance [10], [5], 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-focussed algorithms. On the other hand, social welfare maximizing action profiles, generally, are not individually optimum, thus they are intrinsically less stable than NE.

We aim at linking the welfare maximizing action with the solution of the problem in (3). The next theorem proves that the action selected with highest probability by ODL is the one which solves the global optimization problem.

Theorem 2: Let $\mathcal{G}$ be a K-person game, where each player implements ODL with utility function (7)*, and let it be $\beta > K$. Then, the action profile with the highest social welfare is a solution of* (3).

This theorem states that if the parameter β is greater than the number of clusters K then, the action mostly selected by the algorithm, say $\boldsymbol{a}^*$, satisfies the constraints of the largest possible set of pairs while minimizing the power consumption. The proof of this theorem is given in appendix A.

V. SIMULATION RESULTS

In this section we present the results of the numerical simulations run for showing the performance of ODL in the scenarios introduced in section II.

A. Small Network

Given the model in section II-A, it can be proven that the game $\mathcal{G}$ has one NE composed of the pair $a_1^{NE} = (c_1^{NE}, p_1^{NE})$ and $a_2^{NE} = (c_2^{NE}, p_2^{NE})$ with $c_1^{NE} = 1$, $c_2^{NE} = 2$,$p_1^{NE} = 10$ W and $p_2^{NE} = 0$ W. We denote by $a_1^* = (c_1^*, p_1^*)$ and $a_2^* = (c_2^*, p_2^*)$ the social welfare maximizing action profile with $c_1^* = 2$, $c_2^* = 1$, $p_1^* = 20$ W and $p_2^* = 10$ W. We run 10^5 tests of ODL and trial and error (TE) [1], a NE reaching algorithm. The results are summarized in Table I.

As we can see from the simulation results, ODL achieves a higher social welfare thanks to the higher level of satisfaction in the network.

B. Large network

Here, the network underlying the numerical simulation is the one in II-B. First, we run 10 experiments each of which composed by 6000 iterations of ODL, on a scenario where each cluster is composed by $N = 4$ pairs. The results are summarized in Fig. 3. The upper curve represents the average level of satisfaction, i.e., the average fraction of pairs which are able to satisfy their SINR constraints. The lower curve represents the average amount of power spent in network. Around the 60% of the pairs is able to achieve a $SINR > 10$ dB, while employing an average power of 20 W. In Fig. 4, we

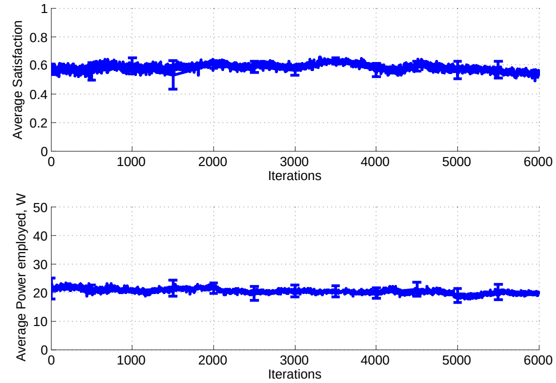

Fig. 3. Performance of the algorithm in a dense network. The upper curve represent the average level of satisfaction in the network, the lower curve the average level of power spent. The standard deviation is plot along with the average.Here, $K = 16$ clusters with 4 pairs share $C = 4$ channels.

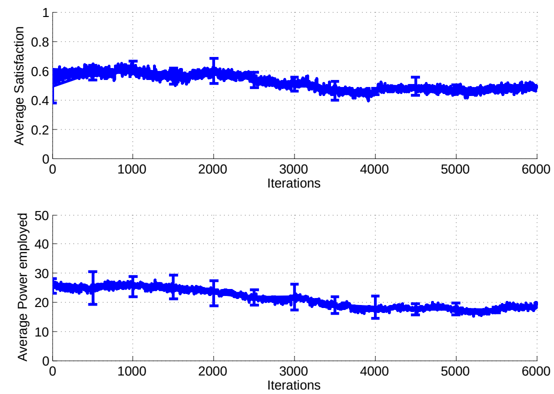

Fig. 4. Performance of the algorithm in a dense network. The upper curve represent the average level of satisfaction in the network, the lower curve the average level of power spent. The standard deviation is plot along with the average.Here, $K = 16$ clusters with 4 pairs share $C = 3$ channels.

reduce the number of available channels to $C = 3$. This reduces the system performance in terms of average satisfaction, due to the fact that the scarcity of spectral resources implies a higher level of mutual interference. The reduction of the level of power employed is a consequence of the utility function chosen. By simple inspection of (7), we can see that if a player cannot be satisfied, for instance because in all the channels the interference level is too high, the power level which maximizes its utility is $p_k = 0$.

C. ODL and TE comparison

In this section, we compare the performance of the TE algorithm (a NE reaching algorithm) introduced in [1], and ODL. Both algorithms share a state machine structure, a stochastic nature and they require the same amount of information. The main difference lies on the converging point. Implementing a social welfare maximizer may come at the cost of stability and of converging time. This can be consider an instance of the exploitation versus exploration trade-off. In the next simulation, we run extensive experiments over the network described in section II-B, where we vary the number of available channels,

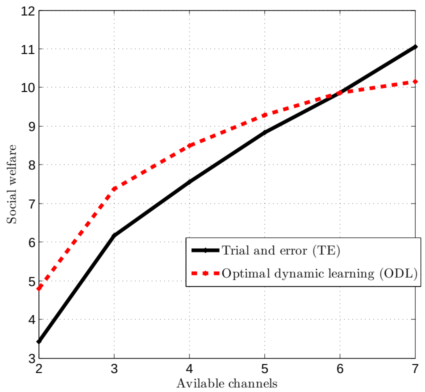

Fig. 5. Comparison between NE-searching algorithm and ODL. The dashed curve represents the average social welfare obtained by ODL, the continuous curve represents the average social welfare obtained by TE both as a function of the available channels.

from $C = 2$ to $C = 7$. The results are represented in Fig. 5. The comparison is performed in terms of social welfare overall the simulation time. The red dashed line represents the 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 well counterbalanced by the gain due to the selection of a well-performing working point.

VI. Conclusions

In this work, we have studied the performance of an algorithm, namely optimal dynamic learning (ODL), suitable for self configuring wireless networks. This algorithm is able to jointly set channel and power level in order to guarantee with high probability that the largest set of devices is able to communicate, while employing the minimal necessary power. ODL requires neither prior information on the network nor any channel state information, and it relies only on local available information. ODL does not focus in implementing a Nash equilibrium (NE), rather it aims at finding a globally optimal solution. Simulation results show that, when the resources of the network are scarce compared to the amount of potential users, this approach brings advantages over NE reaching algorithms.

VII. Acknowledgement

This research work was carried out in the framework of the CORASMA EDA Project B-0781-IAP4-GC.

References

[1] L. Rose, S. M. Perlaza, M. Debbah, and C. L. Martret, "Distributed power allocation with SINR constraints using trial and error learning," in *Proceedings of IEEE Wireless Communications and Networking Conference, WCNC, Paris, France*, Apr. 2012.

[2] H. Iiduka, "Fixed point optimization algorithm and its application to power control in CDMA data networks," *Mathematical Programming*, vol. 1, pp. 1–12, Nov. 2010.

[3] 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.

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

[5] L. Rose, S. M. Perlaza, and M. Debbah, "On the Nash equilibria in decentralized parallel interference channels," in *Proceedings of IEEE Workshop on Game Theory and Resource Allocation for 4G*, Kyoto, Japan, Jun. 2011.

[6] E. Altman, V. Kamble, and H. Kameda, "A Braess type paradox in power control over interference channels," *1st International ICST Workshop on Physics Inspired Paradigms for Wireless Communications and Network*, May 2008.

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

[8] J. R. Marden, H. P. Young, and L. Y. Pao, "Achieving pareto optimality through distributed learning," University of Oxford, Department of Economics, Economics Series Working Papers 557, 2011.

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

[10] 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 Transaction on Signal Processing*, vol. 58, no. 6, pp. 3251–3271, Jun. 2010.

Appendix

Proof: We aim at proving that:

- (i) Let $\boldsymbol{a}^*$ be an action profile which satisfies L^* links and let $\boldsymbol{a}^x$ an action profile which satisfies L^x links, and let $L^* > L^x$, then $W(\boldsymbol{a}^*) > W(\boldsymbol{a}^x)$;
- (ii) Let $\boldsymbol{a}^* = (\boldsymbol{p}^*, \boldsymbol{c}^*)$ be such that $\boldsymbol{a}^* = \arg\max_{\boldsymbol{a}\in\mathcal{A}} W(\boldsymbol{a})$, let L^* be the number of links satisfied with $\boldsymbol{a}^*$ and let $\mathcal{A}^*$ be the set of action profiles which satisfy L^* players, then the power vector $\boldsymbol{p}^*$ is such that $\boldsymbol{p}^* = \arg \min_{\boldsymbol{p}\in\mathcal{P}^K} \sum_k p_k$.

To prove (i), let us write the social welfare as:

$$\begin{aligned} W(\boldsymbol{a}) = & \sum_{k\in\mathcal{K}} \frac{1}{1+\beta N_k} \left(1 - \frac{p_k}{P_{MAX}} + \beta \sum_{i\in\mathcal{N}_k} \mathbb{1}_{\{SINR_i(\boldsymbol{a})>\Gamma\}}\right) \\ = & \frac{\sum_{k\in\mathcal{K}} \left(1 - \frac{p_k}{P_{MAX}}\right) + \beta \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{N}_k} \mathbb{1}_{\{SINR_i(\boldsymbol{a})>\Gamma\}}}{1+\beta N}. \end{aligned}$$

Note that we used the assumption $N_k = N$. Since $0 \leq p_k \leq P_{MAX}$ by definition, we have that:

$$0 \leq \sum_{k\in\mathcal{K}} \left(1 - \frac{p_k}{P_{MAX}}\right) \leq K. \quad (10)$$

By using the left side of inequality (10), and the assumption that $\boldsymbol{a}^*$ is an action profile which satisfies L^* links, we can write $W(\boldsymbol{a}^*) \geq \frac{\beta L^*}{1+\beta N}$. Similarly, using the right side of inequality (10), and the assumption that $\boldsymbol{a}^x$ is an action profile which satisfies L^x links, we can write $W(\boldsymbol{a}^x) \leq \frac{K+\beta L^x}{1+\beta N}$. Since $L^x, L^* \in \mathbb{N}$, we can write the assumption $L^x < L^*$ as $L^x \leq L^* - 1$, which implies $W(\boldsymbol{a}^x) \leq \frac{K+\beta L^*-\beta}{1+\beta N}$. For the assumption that $\beta > K$, we can then write $\frac{K+\beta L^*-\beta}{1+\beta N} < \frac{\beta L^*}{1+\beta N}$. Now, following the chain of inequalities, we can state that $W(\boldsymbol{a}^x) < \frac{\beta L^*}{1+\beta N} \leq W(\boldsymbol{a}^*)$, which proves (i).

To prove (ii), it suffices to note that $W(\boldsymbol{a})$ is monotonic descendent with $\sum_{k\in\mathcal{K}} p_k$, thus, any welfare maximizing action profile must minimize the sum of the power, while keeping constant the number of links satisfied. Which proves (ii) and, thus, our thesis. ■