A *Q*-Learning Game-Theory-Based Algorithm to Improve the Energy Efficiency of a Multiple Relay-Aided Network

Farshad Shams[1], Giacomo Bacci[2,3], and Marco Luise[2,3]*

[1]Dept. Computer Science and Engineering, IMT Institute for Advanced Studies, Lucca, Italy (f.shams@imtlucca.it)
[2]Dip. Ingegneria dell'Informazione, University of Pisa, Pisa, Italy ({giacomo.bacci, marco.luise}@iet.unipi.it)
[3]Consorzio Nazionale Internuniversitario per le Telecomunicazioni (CNIT), Parma, Italy

Abstract

This paper studies a resource allocation problem for a cooperative network with multiple wireless transmitters, multiple full-duplex amplify-and-forward relays, and one destination. A game-theoretic model is used to devise a power control algorithm among all active nodes, wherein the sources aim at maximizing their energy efficiency, and the relays aim at maximizing the network sum-rate. To this end, we formulate a low-complexity *Q*-learning-based algorithm to let the active players converge to the best mixed-strategy Nash equilibrium point, that combines good performance in terms of energy efficiency and overall data rate. Numerical results show that the proposed scheme outperforms Nash bargaining, max-min fairness, and max-rate optimization schemes.

1. Introduction

The ever-increasing demand for high-speed ubiquitous wireless communications calls for efficient solutions in terms of energy expenditure and bandwidth occupation. Among the others, *relay-assisted communication* [1] has become a very promising technique in a number of wireless systems, such as ad-hoc, mesh, and cellular networks. The basic idea is to combine spectral efficiency (SE) and energy efficiency (EE) by transmitting data through several intermediate nodes, called *relays*, that retransmit such data to the receiver, using different schemes: either decode-and-forward (DF), or compress-and-forward (CF), or amplify-and-forward (AF), and either half-duplex or full-duplex [2].

In the literature, there exist many attempts to properly allocate the resources in a relay-aided network. Just to mention a few relevant applications, power control algorithms have been derived for the single-relay (e.g., [3]) and the multiple-relay (e.g., [4]) scenarios. Interactions among the nodes can also be effectively modeled using game theory [5] (e.g., [6-9]). However, most approaches show a relatively high computational complexity, which could seriously undermine their applicability. To address this drawback, we propose a *Q*-learning-based algorithm [10] to achieve the solution of a game that models the power allocation problem, focusing on the full-duplex AF strategy. To the best of our knowledge, power control schemes for a multiple-relay network in which *both* source nodes and relay nodes in a *full-duplex* communication mode are involved in not available in the literature.

The remainder is structured as follows. Section 2 contains the formulation of the problem as a noncooperative game, whose solution is computed in Section 3 using a reinforcement-learning method. Section 4 compares the performance of the proposed algorithm with other methods available in the literature, and Section 5 concludes the paper.

2. Formulation of the resource allocation problem

The uplink of a cooperative relay-aided network, wherein S multiple sources reach the destination through R multiple parallel AF relays working in full-duplex mode, can be modeled as in Fig. 1, where $s, s' \in \mathcal{S} = \{1, \ldots, S\}$ are two generic source nodes, $r \in \mathcal{R} = \{1, \ldots, R\}$ is a generic relay node, g_{ki} denotes the channel gain between transmitter k and receiver i, and $W_i \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) received at node i with power σ^2. Using [11], the sth source's signal-to-interference-plus-noise ratio (SINR) at the destination is

$$\gamma_s = \frac{h_s p_s}{\sigma^2 + \sum_{s' \in \mathcal{S}, s' \neq s} h_{s'} p_{s'}} \tag{1}$$

where p_k, $p_k \leq \bar{p}_k$, is node k's transmit power, with $\bar{p}_k$ denoting node k's maximum power, and

$$h_s = \left(\sqrt{g_{sd}} + \sum_{r \in \mathcal{R}} |\alpha_r| \sqrt{g_{sr} g_{rd}} \right)^2 \Big/ \left(1 + \sum_{r \in \mathcal{R}} |\alpha_r|^2 g_{rd} \right) \tag{2}$$

with α_r denoting the scaling factor adopted by relay r to implement the AF strategy in the AWGN scenario, with

$$|\alpha_r| = \sqrt{p_r \Big/ \left(\sigma^2 + \sum_{s \in \mathcal{S}} h_{sn} p_s \right)}. \tag{3}$$

978-1-4673-5225-3/14/$31.00 ©2014 IEEE

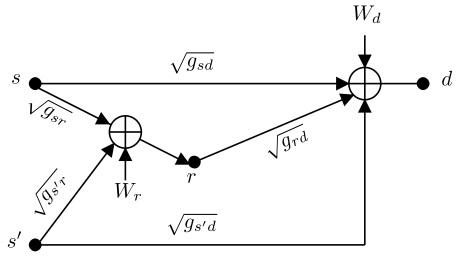


Figure 1 – Relay-aided source-to-destination communication.

Using (1), we can express the Shannon capacity achievable by source nose s as $c_s=\Delta f\cdot\log_2(1+\gamma_s)$ [b/s], where Δf denotes the signal bandwidth. Hence, when inspecting (1)–(3), it is straightforward to note the coupling among the powers of all nodes (sources and relays) on the rate achievable by each source node. A joint power control, that consists in finding an optimal vector of transmit powers $\mathbf{p}=[p_1, \dots, p_K]$, where p_k is such that $0 \le p_k \le \bar{p}_k$ for all $k \in \mathcal{S} \cup \mathcal{R}$, such that the network performance can be increased, also including QoS constraints on each source node s's minimum rate θ_s, is thus highly desirable, and at the same time quite challenging. To investigate the solution to this problem, we will use the analytical tools of game theory [5], whose aim is to help us predict the behavior of rational agents with conflicting interests. To this end, we model the interaction among different nodes as the following noncooperative game:

$$\mathrm{G} = \left\{ \mathrm{K}, \left\{ \mathrm{P}_k \right\}_{k \in \mathrm{K}}, \left\{ u_k \left(p_k ; \mathbf{p}_{\backslash k} \right) \right\}_{k \in \mathrm{K}} \right\} \tag{4}$$

where:

- $\mathcal{K} = \mathcal{S} \cup \mathcal{R}$, with $|\mathcal{K}| = K = S + R$, is the set of all active nodes (the *players* of the game);
- $\mathcal{P}_k$ is the *discrete* set of user k's transmit power, defined as $\mathcal{P}_k = \{0, \Delta p_k, 2\,\Delta p_k, \dots, L\Delta p_k\}$, where $L + 1$, with $1 \le L < \infty$, denotes the number of power levels (including zero power), and $\Delta p_k = \bar{p}_k / L$ is the power step; and
- $u_k(p_k;\mathbf{p}_{\backslash k})$ is the *utility function* of each user $k \in \mathcal{K}$, where $\mathbf{p}_{\backslash k} = \mathbf{p} \setminus p_k$ is the power vector of all nodes (including both sources and relays) excluding source k's power p_k, defined as

$$u_k \left(p_k ; \mathbf{p}_{\backslash k} \right) = \begin{cases} \dfrac{c_k \left(\mathbf{p} \right)}{p_k + \Psi} & \text{s.t. } c_k \left(\mathbf{p} \right) \ge \theta_k \qquad \text{if } k \in \mathrm{S} \\ \displaystyle\sum_{s \in \mathrm{S}} \frac{c_s \left(\mathbf{p} \right)}{\Psi} & \text{s.t. } c_s \left(\mathbf{p} \right) \ge \theta_s \ \forall s \in \mathrm{S} \quad \text{if } k \in \mathrm{R} \end{cases} \tag{5}$$

  to account for the different needs demanded by the two classes of users $\mathcal{S}$ and $\mathcal{R}$. The goal of each source $k \in \mathcal{S}$ is to trade off its achievable channel capacity $c_k(\mathbf{p})$ (3) with its power consumption $p_k+\Psi$, where $\Psi>0$ is the (nonradiative) circuit power [12, 13], thus increasing its EE; and the goal of each relay $k \in \mathcal{R}$ is to increase the network sum-rate, that depends on each source s's $c_s(\mathbf{p})$, that is in turn a function of p_k according to (1)-(3).

A close inspection of the utilities (5) reveals that including the QoS constraints θ_s introduces a coupling between the power sets for all players $k \in \mathcal{K}$. Moreover, note that, as the number of players K is finite, and the number of actions available to each player $L + 1$ is also finite, $\mathcal{G}$ is called a finite game [5]. To solve the maximization problem

$$p_k^* = \arg \max_{p_k \in \mathrm{P}_k} u_k \left(p_k ; \mathbf{p}_{\backslash k} \right) \tag{6}$$

in a scalable and distributed way, and thus keeping its complexity low, we can make use of the analytical tools of noncooperative game theory [5]. Solutions to (6) are represented by mixed-strategy Nash equilibria, defined as follows.

Definition 1: A mixed-strategy Nash equilibrium for a game $\mathcal{G}$ is a K-tuple of vectors $[\boldsymbol{\sigma}_1^*, \dots, \boldsymbol{\sigma}_K^*]$, with $\boldsymbol{\sigma}_k^* \in [0,1]^{L+1}$, such that, for all $k \in \mathcal{K}$ and all $\boldsymbol{\sigma}_k \in [0,1]^{L+1}$,

$$\sum_{p_k \in \mathrm{P}_k} \sum_{\mathbf{p}_{\backslash k} \in \mathbf{P}_{\backslash k}} \boldsymbol{\sigma}_{\backslash k}^* \left(\mathbf{p}_{\backslash k} \right) \sigma_k^* \left(p_k \right) u_k \left(p_k ; \mathbf{p}_{\backslash k} \right) \ge \sum_{p_k \in \mathrm{P}_k} \sum_{\mathbf{p}_{\backslash k} \in \mathbf{P}_{\backslash k}} \boldsymbol{\sigma}_{\backslash k}^* \left(\mathbf{p}_{\backslash k} \right) \sigma_k \left(p_k \right) u_k \left(p_k ; \mathbf{p}_{\backslash k} \right) \tag{7}$$

where $p_k \in \mathcal{P}_k$ is a pure strategy, $\boldsymbol{\mathcal{P}}_{\backslash k} = \times_{i \ne m} \mathcal{P}_i$ is the cartesian product of all strategy sets other than k's one, and, likewise, $\boldsymbol{\sigma}_{\backslash k}^*(\mathbf{p}_{\backslash k}) = \Pi_{\,i \ne m}\, \sigma_i^*(p_i)$, where the product stems from the independence of each player's action with respect to the others'.

Theorem 1([14]): In every finite static game $\mathcal{G}$ there exists at least one mixed-strategy Nash equilibrium.

3. *Q*-learning-based algorithm

Once the existence of (at least) one mixed-strategy Nash equilibrium in $\mathcal{G}$ is assessed, we now aim at computing it. The algorithm proposed in this paper, which adapts the one derived in [15], is based on reinforcement-learning techniques [10], and runs as follows:

Initialization:

`for` every $k \in \mathcal{K}$ `do`

`for` every $\mathbf{p} \in \boldsymbol{\mathcal{P}} = \mathcal{P}_k \times \boldsymbol{\mathcal{P}}_{\setminus k}$ `do`

set $Q_k^0(\mathbf{p}) = u_k(\mathbf{p})$

`end for`

`end for`

Feasibility check:

`if` $\left(\exists k \in \mathrm{K} \text{ s.t. } Q_k^0(\mathbf{p}) = 0 \forall \mathbf{p} \in \mathbf{P}\right)$ `then` **exit** `else` **set** t=0 and a tolerance ε=1;

Loop:

`repeat`

`for` every $k \in \mathcal{K}$ `do`

`for` every $\mathbf{p} \in \boldsymbol{\mathcal{P}} = \mathcal{P}_k \times \boldsymbol{\mathcal{P}}_{\setminus k}$ `do`

update $\pi_k^t(\mathbf{p}) = \dfrac{\exp\left\{\sum_{l=0}^{t} (\delta_k)^l \cdot u_k(\mathbf{p}) / T_k(t)\right\}}{\sum_{\mathrm{P}_k} \sum_{\mathbf{P}_{\setminus k}} \exp\left\{\sum_{l=0}^{t} (\delta_k)^l \cdot u_k(\mathbf{p}) / T_k(t)\right\}}$, where $\delta_k \in (0,1)$, and $T_k(t)$ is a temperature function;

`end for`

compute $\mathbf{p}_k^{(t)} = \arg\max_{\mathbf{p} \in \mathbf{P}} \pi_k^t(\mathbf{p})$

`end for`

`for` every $k \in \mathcal{K}$ `do`

update $Q_k^{t+1}(\mathbf{p}_k^{(t)}) \leftarrow (1 - f_k(t+1)) \cdot Q_k^t(\mathbf{p}_k^{(t)}) + f_k(t+1) \cdot \left(u_k(\mathbf{p}_k^{(t)}) + \delta_k \cdot Q_k^t(\mathbf{p}_k^{(t)}) \cdot \prod_{i=1}^{K} \pi_i^t(\mathbf{p}_k^{(t)})\right)$, with learning rate $f_k(t)$;

`end for`

update $t = t + 1$;

`until` $\max_{k \in \mathrm{K}} \left| Q_k^t(\mathbf{p}_k^{(t)}) - Q_k^{t-1}(\mathbf{p}_k^{(t)}) \right| \le \varepsilon$

Output: **compute** the mixed-strategy Nash equilibrium, with elements $\sigma_k^*(p_k) = \sum_{i \in \mathrm{K}, i \neq k} \sum_{p_i \in \mathrm{P}_i} \pi_k^t(p_k; \mathbf{p}_{\setminus k})$.

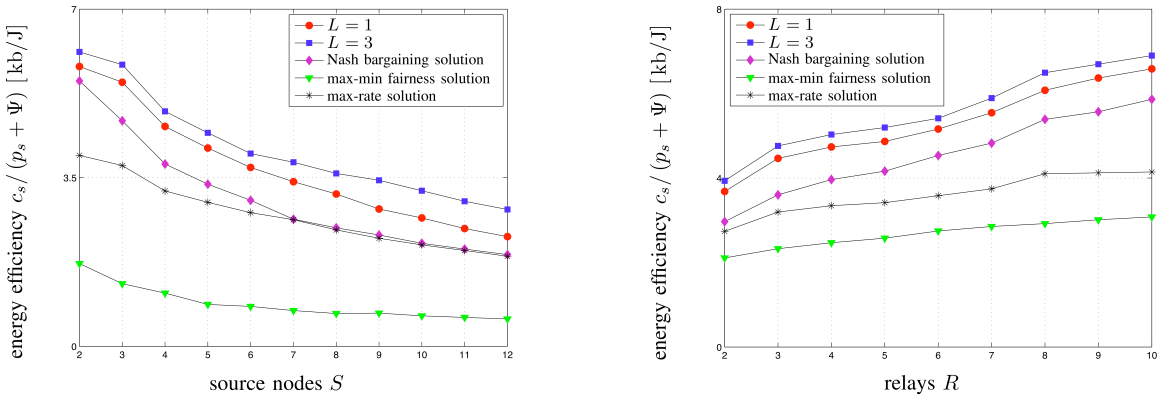
4. Numerical results

In this section, we show the performance of the proposed algorithm to control the power in a multiple-relay-aided communication network scenario, and compare it with that of well-known power allocation schemes, namely:

$$\begin{aligned} &\text{Nash bargaining solution (NBS):} && \max_{p_k \in [0, \bar{p}_k]} \prod_{s \in \mathcal{S}} \left(c_s(\mathbf{p}) - \theta_s\right) / \left(p_s + \Psi\right) \\ &\text{max-min fairness:} && \max_{p_k \in [0, \bar{p}_k]} \min_{s \in \mathcal{S}} c_s(\mathbf{p}) / \left(p_s + \Psi\right) \\ &\text{max-rate solution:} && \max_{p_k \in [0, \bar{p}_k]} \sum_{s \in \mathcal{S}} c_s(\mathbf{p}) \end{aligned} \tag{8}$$

Throughout the simulations, we use Δf = 10.938 kHz, σ^2 = 10 nW, Ψ = 100 mW, $\bar{p}_k$ = 1W for all $k \in \mathcal{K}$, and θ_s=1 kb/s for all source nodes $s \in \mathcal{S}$. The distances of relays and source nodes from the destination are uniformly distributed between 10 and 100 m, and a 24-tap channel model is used to reproduce the effects of shadowing. We also set $f_k(t) = t^{-0.8}$, δ_k = 0.85, and $T_k(t) = 10^{-2}(\max_{\mathbf{p} \in \boldsymbol{\mathcal{P}}} u_k(\mathbf{p})) \cdot \exp\{-10^{-2} t (\max_{\mathbf{p} \in \boldsymbol{\mathcal{P}}} u_k(\mathbf{p}))\}$, that provide a good tradeoff between EE and SE, based on an exhaustive search [11], not reported here for the sake of brevity. To reduce the computational burden, which is exponentially increasing with the number of power steps L+1, we select the cases L = 1, corresponding to the situation in which the sum-rate is maximized and the computational load is the minimum one, at the cost of a reduced EE, and L = 3, which provides an interesting tradeoff between SE, EE, and computational complexity of the algorithm. Note that, using such values of L, the proposed algorithm converges after a few iteration steps (typically, ≤ 3) [11].

We will compare the performance of our proposed algorithm, using L = 1 (circles) and L = 3 (squares), with NBS (diamonds), max-min fairness (lower triangles), and max-rate (asterisks) solutions. Figs. 2 and 3 report the average EE as functions of the number of source nodes S, and the number of relays R, respectively. As expected, the EE is decreasing with S and increasing with R. When R is fixed, increasing S increases the multiple access interference, thus reducing the EE. On the contrary, increasing R while S is constant increases the sum-rate in the long run, without additional power expenditure at the source side. Note that the case L = 3 outperforms the case L = 1. However, even in the extreme case L = 1 (i.e., each node selects either zero power or its maximum one), the proposed algorithm outperforms the well-known solutions (8). Similar conclusions can be drawn for the average SE, reported in [11], which shows that the proposed algorithm achieves higher sum-rates than NBS and max-min fairness, while paying an acceptable performance gap with respect to the max-rate criterion.



(a) (b)

Figure 2 – Average source node's EE as a function of: (a) the number of sources *S* (with *R*=4); and (b) the number of relays *R* (with *S*=4).

5. Conclusion

In this work, we derived a *Q*-learning-based algorithm to address the power control for a network populated by multiple sources, multiple parallel relays, and one destination. To this aim, we modeled each active node as a player in a non-cooperative finite game: each source seeks to maximize its own energy efficiency; and each relay seeks to maximize the network sum-rate. Simulation results show that the proposed low-complexity algorithm outperforms the Nash bargaining solution and the max-min fairness approach in terms of both energy efficiency and network sum-rate, and significantly outperforms the max-rate solution in terms of energy efficiency, while paying a tolerable performance gap in terms of network sum-rate. Further work is needed to assess the feasibility of the problem given a particular network realization, and to extend the formulation of the problem to a multicarrier system.

6. Acknowledgments

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the EU's FP7 under REA Grant agreement no. PIOF-GA-2011-302520 GRAND-CRU, and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM# (GA no. 318306).

7. References

1. A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.
2. G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Information Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.
3. Z. Sahinoglu and P. Orlik, "Regenerator versus simple-relay with optimum transmit power control for error propagation," *IEEE Commun. Letters*, vol. 7, no. 9, pp. 416–418, Sept. 2003.
4. J. Paredes and A. Gershman, "Relay network beamforming and power control using maximization of mutual information," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4356–4365, Dec. 2011.
5. M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA: MIT Press, 1994.
6. J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 26, no. 7, pp. 1226–1237, Sept. 2008.
7. S. Ren and M. van der Schaar, "Pricing and distributed power control in wireless relay networks," *IEEE Trans. Signal Processing*, vol. 59, no. 6, pp. 2913–2926, June 2011.
8. H. Khayatian, R. Saadat, and J. Abouei, "Coalition-based approaches for joint power control and relay selection in cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 835–842, Feb. 2013.
9. I. Stupia, L. Vandendorpe, L. Sanguinetti, and G. Bacci, "Distributed energy-efficient power optimization for relay-aided heterogeneous networks," in *Intl. Workshop Wireless Networks: Communication, Cooperation and Competition*, Hammamet, Tunisia, May 2014, submitted.
10. C. J. Watkins and P. Dayan, "Q-Learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.
11. F. Shams, G. Bacci, and M. Luise, "Game-theoretic power control for multiple-relay cooperative networks," *IEEE Trans. Wireless Commun.*, 2014, in preparation.
12. G. Miao, N. Himayat, and G. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.
13. C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.
14. J. F. Nash, "Equilibrium points in n-person games," *Proc. National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, Jan. 1950.
15. J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *J. Machine Learning Research*, vol. 4, pp. 1039–1069, 2003.