Enhancing Spectral Efficiency in Dense Wireless Networks through Learning

Universitat Pompeu Fabra Barcelona

Francesc Wilhelmi

Supervisor UNAM: J. Gómez
Supervisors UPF: B. Bellalta, A. Jonsson, C. Cano

31/08/18

Table of contents

Outline

Problem description

Spatial Reuse (SR) enhancement in dense Wireless Networks

- Transmit Power Control (TPC)
- Carrier Sense Threshold (CST) adjustment
- Dynamic Channel Selection (DCA)

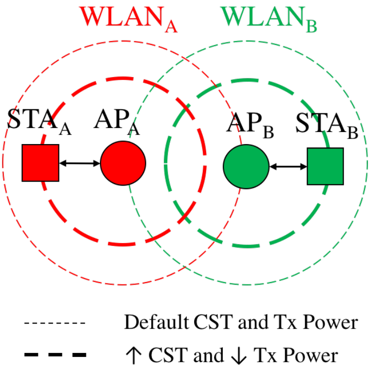


Figure 1: SR enhancement through TPC and CST adjustment

Context - Use case

Dense IEEE 802.11 WLANs

- Unplanned (chaotic deployments)
- Decentralized (local information only)
- Unpredictable interactions between overlapping networks

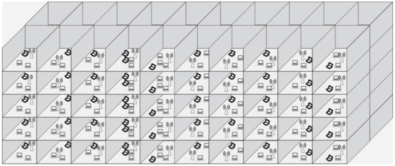


Figure 2: TGax residential scenario. Image retrieved from [1].

Potential Solution

Online Learning

1. Uncertainty → **no** information exchange
2. Adversarial setting → the reward is influenced by the **environment**
3. Complexity and delay-sensitive → need to find an **approximation** of the optimal solution, rather than computing it

Action	**Effect**			
	Parallel Transmissions	Data Rate	Collisions probability (by hidden node)	Energy Consumption
↑ Power	↓	↑	↓	↑
↓ Power	↑	↓	↑	↓
↑ CST	↑	-	↑	↑*
↓ CST	↓	-	↓	↓*

Table 1: Effects of TPC and CST adjustment

Outline

Related Work

Surveys

- Cognitive radio [1]
- Wireless Sensor Networks (WSN) [2, 3]
- Ad-hoc networks [4]

Related to this problem

- Q-learning for channel selection [5-8] and power adjustment [9, 10]
- MABs to Power control in D2D networks [11, 12]
- MABs to DCA & TPC [13]
- Structured MABs for combinatorial optimization problems [14, 15]
- MABs for decentralized channel access [16, 17]

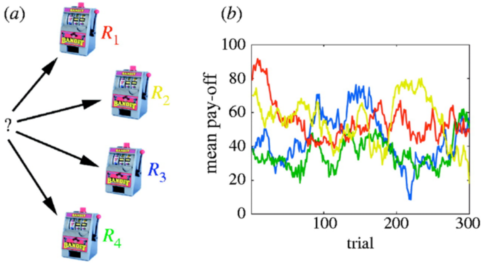
Outline

The Multi-Armed Bandit problem

Formal definition

A game in which the following steps are repeated in $t = 1, 2, \ldots, T$:

1. The environment fixes an assignment of rewards $r_{a,t}$ for each action $a \in [K] \stackrel{\text{def}}{=} \{1, 2, \ldots, K\}$,
2. the learner chooses action $a_t \in [K]$,
3. the learner obtains and observes reward $r_{a_t,t}$



MABs application into Decentralized WLANs

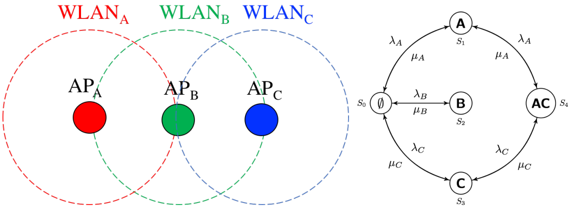
Use case

- Adversarial setting (N WLANs make actions simultaneously)
- Actions consist in {Channel, Transmit Power, CST} combinations
- The reward is set as a function of the WLANs performance
  - Throughput
  - Delay
  - Packets sent vs packets lost
  - ...

Outline

Analytical Models

- Continuous Time Markov Networks (CTMNs) [18]
  - Matlab simulator for building CTMNs: `https://github.com/sergiobarra/SFCTMN`



$$\Gamma_i = L \sum_{s \in S} \mu_{n_{i,s}} \pi_s$$

(a) Scenario (b) CTMN

- Bianchi's model for throughput calculation based on the Distributed Coordination Function (DCF) [19]

Simulation Tools

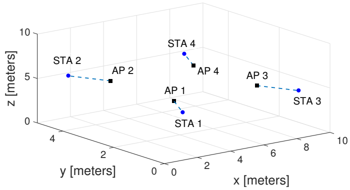
IEEE 802.11ax-based network simulator

- Open source: `https://github.com/wn-upf/Komondor`
- Current tools lack of flexibility for including novel mechanisms in WLANs
- Programmed to include ML-based agents

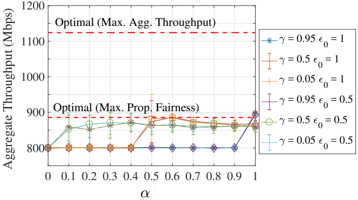
Outline

Conference paper

- Presented at PIMRC'17 (Montreal) [20]
- **Goal:** to improve spectral efficiency in adversarial WNs
- **Method:** a stateless version of Q-learning
- **Conclusions:**
  - Close-to-optimal solutions can be achieved
  - Competitiveness involves the non-existence of a Nash Equilibrium
  - Strong throughput fluctuations



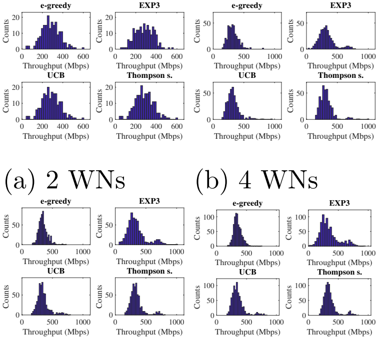
(a) Scenario



(b) Results (100 repetitions of 1,000 iterations)

Journal paper (I)

- Submitted to Ad-hoc Networks [21]
- **Goal:** to improve spectral efficiency in adversarial WNs
- **Method:** Multi-Armed Bandits (several algorithms)
- **Conclusions:**
  - Collaborative behavior even if acting selfishly
  - Trade-off between variability and performance
  - Poor performance in ε-greedy and EXP3
  - Good results in UCB and Thompson sampling



(a) 2 WNs (b) 4 WNs

(c) 6 WNs (c) 8 WNs

Figure 5: Histogram of the avg. throughput (100 repetitions of 10,000 iterations)

Journal paper (II)

- Submitted to JNCA [22]
- **Goal:** to show the major challenges and opportunities of applying ML to the SR problem
- **Method:** MABs
- **Conclusions:**
  - Importance of designing the reward
  - Selfish vs Environment-aware rewards
  - Learning in presence of asymmetries

Ongoing work

Thesis

- Overview of IEEE 802.11ax Spatial Reuse
- Learning Spatial Reuse in IEEE 802.11ax WLANs
- ML-based architecture for WLANs
    - Rough idea: flexible architecture able to offload tasks to different layers
    - Focus on the existing mechanisms to aid the ML-based operation

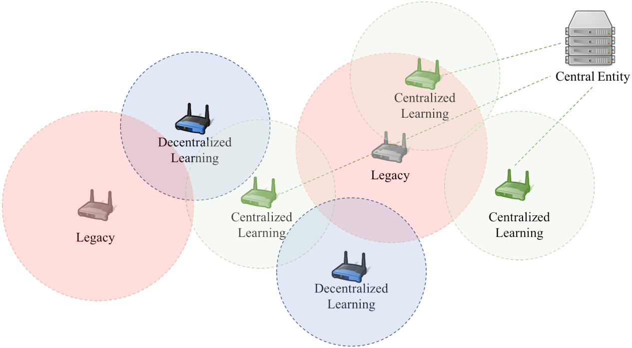
Other projects

- Contributions to FG-ML5G (ITU)
- Fon

Future Work

Candidate research lines

- Improved decentralized learning: inference
- Distributed learning
- Centralized learning



Any questions?

?

Francesc Wilhelmi
francisco.wilhelmi@upf.edu
PhD student
Department of Communication and Information Technologies
Universitat Pompeu Fabra (Barcelona)

References

[1] Afaqui, M. Shahwaiz, et al. "Evaluation of dynamic sensitivity control algorithm for IEEE 802.11 ax." Wireless Communications and Networking Conference (WCNC), 2015 IEEE. IEEE, 2015.

[2] BKASSINY, Mario; LI, Yang; JAYAWEERA, Sudharman K. A survey on machine-learning techniques in cognitive radios. IEEE Communications Surveys & Tutorials, 2013, vol. 15, no 3, p. 1136-1159.

[3] ALSHEIKH, Mohammad Abu, et al. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. IEEE Communications Surveys & Tutorials, 2014, vol. 16, no 4, p. 1996-2018.

[4] DI, Ma; JOO, Er Meng. A survey of machine learning in wireless sensor netoworks from networking and application perspectives. En Information, Communications & Signal Processing, 2007 6th International Conference on. IEEE, 2007. p. 1-5.

[5] FORSTER, Anna. Machine learning techniques applied to wireless ad-hoc networks: Guide and survey. En Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on. IEEE, 2007. p. 365-370.

[6] Nie, J., & Haykin, S. (1999). A Q-learning-based dynamic channel assignment technique for mobile communication systems. IEEE Transactions on Vehicular Technology, 48(5), 1676-1687.

[7] Li, H. (2009, October). Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case. In Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on (pp. 1893-1898). IEEE.

[8] Sallent, O., Pérez-Romero, J., Ferrús, R., & Agustí, R. (2015, June). Learning-based coexistence for LTE operation in unlicensed bands. In Communication Workshop (ICCW), 2015 IEEE International Conference on (pp. 2307-2313). IEEE.

[9] Rupasinghe, N., & Güvenç, İ. (2015, March). Reinforcement learning for licensed-assisted access of LTE in the unlicensed spectrum. In Wireless Communications and Networking Conference (WCNC), 2015 IEEE (pp. 1279-1284). IEEE.

References

[10] Bennis, M., & Niyato, D. (2010, December). A Q-learning based approach to interference avoidance in self-organized femtocell networks. In GLOBECOM Workshops (GC Wkshps), 2010 IEEE (pp. 706-710). IEEE.

[11] Bennis, M., Guruacharya, S., & Niyato, D. (2011, December). Distributed learning strategies for interference mitigation in femtocell networks. In Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE (pp. 1-5). IEEE.

[12] Maghsudi, S., & Stańczak, S. (2015). Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. IEEE Transactions on Vehicular Technology, 64(10), 4565-4578.

[13] Maghsudi, S., & Stańczak, S. (2015). Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting. IEEE Transactions on Wireless Communications, 14(3), 1309-1322.

[14] Wilhelmi, F., Cano, C., Neu, G., Bellalta, B., Jonsson, A., & Barrachina-Muñoz, S. (2017). Collaborative Spatial Reuse in Wireless Networks via Selfish Multi-Armed Bandits. arXiv preprint arXiv:1710.11403.

[15] Gai, Y., Krishnamachari, B., & Jain, R. (2012). Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. IEEE/ACM Transactions on Networking (TON), 20(5), 1466-1478.

[16] Combes, R., & Proutiere, A. (2015). Dynamic rate and channel selection in cognitive radio systems. IEEE Journal on Selected Areas in Communications, 33(5), 910-921.

References

[17] Liu, K., & Zhao, Q. (2010). Distributed learning in multi-armed bandit with multiple players. IEEE Transactions on Signal Processing, 58(11), 5667-5681.

[18] Bellalta, B., Zocca, A., Cano, C., Checco, A., Barcelo, J., & Vinel, A. (2014). Throughput analysis in CSMA/CA networks using continuous time Markov networks: a tutorial. In Wireless Networking for Moving Objects (pp. 115-133). Springer, Cham.

[19] Bianchi, G. (2000). Performance analysis of the IEEE 802.11 distributed coordination function. IEEE Journal on selected areas in communications, 18(3), 535-547.

[20] Wilhelmi, F., Bellalta, B., Cano, C., & Jonsson, A. (2017, October). Implications of decentralized Q-learning resource allocation in wireless networks. In Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on (pp. 1-5). IEEE.

[21] Wilhelmi, F., Cano, C., Neu, G., Bellalta, B., Jonsson, A., & Barrachina-Muñoz, S. (2017). Collaborative Spatial Reuse in Wireless Networks via Selfish Multi-Armed Bandits. arXiv preprint arXiv:1710.11403.

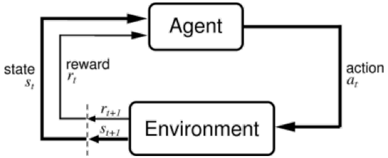
[22] Wilhelmi, F., Barrachina-Muñoz, S., Cano, C., Bellalta, B., Jonsson, A., & Neu, G. (2018). Potential and Pitfalls of Multi-Armed Bandits for Decentralized Spatial Reuse in WLANs. arXiv preprint arXiv:1805.11083.

Backup: Reinforcement Learning

Goal

An agent attempts to learn a policy given the observations it does. The goal is to maximize the expected future cumulative reward.

- No supervisor (only reward signal)
- Delayed feedback & sequentiality
- Actions affect the environment



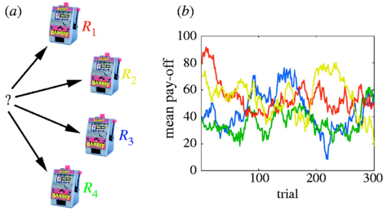
$\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}\}$

- $\mathcal{S}$: set of states
- $\mathcal{A}$: set of actions
- $\mathcal{R}$: set of rewards
- $\mathcal{T}$: transitions probabilities

Backup: Multi-Armed Bandits

Frames the exploration/exploitation trade-off. The hidden reward distributions must be learned while maximizing the gains.

- Action-selection strategies to cope with hidden distributions (ε-greedy, EXP3, UCB...)
- Several variants (contextual, adversarial, stochastic, restless...)
- States-independent
- Reward becomes regret: $R_n = \sum_{t=1}^{n} l_{t,I_t} - \min_{i \in K} \sum_{t=1}^{n} l_{t,i}$



Backup: Thompson sampling

Thompson sampling [19] is a Bayesian action-selection technique

- It constructs a probabilistic model of the rewards and assumes a prior distribution of the parameters of said model
- Keeps track of the posterior distribution of the rewards, and pulls arms randomly in a way that the drawing probability of each arm matches the probability of the particular arm being optimal
- For the sake of practicality, we aim to apply Thompson sampling using a Gaussian model for the rewards with a standard Gaussian prior as suggested in [20].
- In adversarial wireless networks, it has been shown to perform better than using the magnitude of the reward [9]

Backup: Applied Thompson sampling

Algorithm 1: Implementation of Multi-Armed Bandits (Thompson sampling) in a WN

Function Thompson Sampling (SNR, $\mathcal{A}$);

Input : SNR: information about the Signal-to-Noise Ratio received at the STA
$\mathcal{A}$: set of possible actions in $\{a_1, ..., a_K\}$

initialize: $t = 0$, for each arm $a_k \in \mathcal{A}$, set $\hat{r}_k = 0$ and $n_k = 0$

while *active* **do**

- For each arm $a_k \in \mathcal{A}$, sample $\theta_k(t)$ from normal distribution $\mathcal{N}(\hat{r}_k, \frac{1}{n_k+1})$
- Play arm $a_k = \underset{k=1,...,K}{\text{argmax}}\ \theta_k(t)$
- Observe the throughput experienced Γ_t
- Compute the reward $r_{k,t} = \frac{\Gamma_t}{\Gamma^*}$, where $\Gamma^* = B \log_2(1 + \text{SNR})$
- $\hat{r}_{k,t} \leftarrow \frac{\hat{r}_{k,t} n_{k,t} + r_{k,t}}{n_{k,t}+2}$
- $n_{k,t} \leftarrow n_{k,t} + 1$
- $t \leftarrow t + 1$

end