Response Letter of "Potential and Pitfalls of Multi-Armed Bandits for Decentralized Spatial Reuse in WLANs"

Journal on Network and Computer Applications

Francesc Wilhelmi, Sergio Barrachina-Muñoz, Boris Bellalta, Cristina-Cano,
Anders-Jonsson & Gergely-Neu
francisco.wilhelmi@upf.edu

Dept. of Information and Communication Technologies
Universitat Pompeu Fabra (UPF), Barcelona

This manuscript is a revised version of the manuscript with id JNCA-D-18-00769. We would like to thank the reviewers for their comments, which have allowed us to improve our submission, as well as the editor for allowing us to revise our work for publication in this journal. We have performed a thorough revision of the paper to address all the concerns raised by the reviewers. In this letter we address the comments from the reviewers and point out the changes in the revised version of our manuscript (which are highlighted in blue to facilitate revision). In addition, we improved some parts regarding writing and readability (highlighted in orange). Before providing the specific response to each one of the reviewers' comments, we next summarize the main changes introduced in the new version of the paper:

- We have now further motivated the utilization of Multi-Armed Bandits (MABs) to address the decentralized Spatial Reuse (SR) problem in IEEE 802.11 Wireless Local Area Networks (WLANs).
- We have provided some parallelisms between the considerations of learning in a decentralized way and the algorithms proposed in this paper.
- We elaborated a new subsection in which we provide some insights on learning in dynamic environments.
- We have analyzed the performance of Thompson sampling in front of the well-known and popular action-selection strategy ε-greedy.
- Some notes are given in the conclusions with regard to the feasibility and potential of including beamforming to the SR problem.
- We corrected spelling and other writing errors.
- We reprocessed some parts of the text with the aim of facilitating its readability.

We hope the changes made in the revised version of this manuscript provide a clarification on the issues previously raised.

With best regards,

Francesc Wilhelmi, Sergio Barrachina-Muñoz, Boris Bellalta, Cristina-Cano,
Anders-Jonsson & Gergely-Neu
Barcelona (Spain), October 5, 2018

1 Reviewer #1

Comment R1.1

Real-life scenarios are dynamic. This problem is, in my opinion, only vaguely tackled in the paper. Some details about how the proposed algorithms may tackle this problem, including their limitations in this context are needed.

Response. We strongly agree with this statement, and concur that considering network dynamics is essential in order to assess the actual performance that can be achieved when applying Reinforcement Learning (RL) to wireless networks. We added Subsection *4.2.3. Learning in Dynamic WLANs* to the manuscript, with the aim of shedding some light to the matter. This subsection provides a description on how dynamic scenarios can be addressed. In addition, a use case of a dynamic scenario is presented, in which the performance of the proposed algorithm is shown.

We hope that this new contribution encourages other researchers to delve into the topic. In particular, we left as future work the thorough analysis of applying learning in dynamic environments. To that purpose, a preliminary and fundamental requirement consists in exhaustively analyzing the performance of dynamic scenarios with multiple overlapping WLANs. Analyzing the throughput experienced in those environments would allow us to better understand the dynamics of the rewards.

Comment R1.2

The authors suggest that the proposed algorithms may be used inside a decentralized mechanism. The authors discuss this issue in subsection 4.2 but this discussion is general (not related to the two new algorithms). They must be more precise in the case of the two algorithms (by adding a small paragraph inside 4.2 to address this issue).

Response. We modified Subsection *4.2. Considerations of Decentralized Learning in WLANs* to link the current explanations with the two proposed algorithms. In particular, we discuss the implications of applying each algorithm to decentralized scenarios. Such a discussion is later extended in the results section *5. Performance Evaluation*.

Comment R1.3

WLAN range used in simulation seems unrealistic. In some figures the WLAN range seems to be less than 6-7m (see Fig. 7 a for example).

Response. We agree on the fact that the propagation losses may seem to be very strong. However, we took as reference the residential scenario recommended in the IEEE 802.11ax amendment (**?**), and which can be found in 11-14-0980 TGax Simulation Scenarios. Such a scenario frames a residential building with strong path-loss effects due to the high number of obstacles in terms of walls and floors. In particular, since we refer to random scenarios, we did not reproduce the exact scenario provided by the 11ax standard. Instead, we

have captured its essence by proposing a model that takes into account the walls and floor frequencies, rather than the actual location of walls and floors. The formula that we have used for the incurred loss of a given data transmission, which appears in the manuscript's appendix, is:

$$\mathrm{PL}_d = 40.05 + 20\log_{10}\left(\frac{f_c}{2.4}\right) + 20\log_{10}(\min(d,5)) + I_{d>5}\cdot 35\log_{10}\left(\frac{d}{5}\right) + 18.3F^{\frac{F+2}{F+1}-0.46} + 5W,$$

where f_c is the central frequency in GHz, d is the distance between the transmitter and the receiver in meters, and F and W are the average number of floors and walls traversed per meter, respectively. Note that F and W have been obtained from the actual number of walls and floors of the 11ax residential scenario. In particular, the scenario has the following characteristics:

- 5 floors, 3 m height in each floor
- 2x10 apartments in each floor
- Apartment size: 10m x 10m x 3m

With the aim of clarifying this point to the reader, we added further details on the path-loss model in Appendix *A. Wireless Environment*.

Comment R1.4

The abbreviations AP and STA are not defined the first time they appear in the text.

Response. We have corrected this in the manuscript, as indicated by Reviewer 1.

Comment R1.5

There are few typos and incorrectly shaped sentences. For example in the middle of page 13 we have the following incorrect sentence "Furthermore, due to the lack of coordination between WLANs, the abovementioned learning procedure would done in a disorganized way." and also, in the following sentence ("Accordingly, from a global network perspective ..."), it is written "them" instead of "they".

Response. We have performed a throughout revision of the article, following this comment from Reviewer 1.

Comment R1.6

The legend of Fig.10b must be moved outside the figure.

Response. We have corrected this in the manuscript, as indicated by Reviewer 1.

Comment R1.7

On page 25, inside subsection 5.2 it is written: "WLANs are uniformly distributed at random in the scenario...", which is unclear. Probably it should be "WLANs are uniformly randomly distributed in the scenario...", Please rectify.

Response. We have corrected this in the manuscript, as indicated by Reviewer 1.

2 Reviewer #2

Comment R2.1

The authors should provide more justification for using MABs to address the SR problem.

Response. We further justified the utilization of MABs for the decentralized SR problem at the beginning of Section *4. Multi-Armed Bandits for Decentralized Spatial Reuse*. In particular, we show that MABs have been previously used to address the exploration-exploitation trade-off, for scenarios with a high degree of uncertainty. In addition, we find parallelisms between scenarios with uncertainty and the decentralized SR problem.

Comment R2.2

The theoretical data rate provided by the maximum MCS is not clear and unjustified.

Response. We added additional information to footnote 9 in Subsection *4.2.1 Reward Definition*. However, we would also like to provide further details in this response letter.

In this regard, we have provided the data rate corresponding to the maximum MCS permitted by the IEEE 802.11ax for single transmissions. Accordingly, such a data rate is obtained from the necessary time to carry out the entire transmission, including overheads, which can be obtained by emulating the operation of the Distributed Coordination Function (DCF). We have clarified this point in the manuscript.

In particular, the data rate of a given data transmission is obtained as follows:

$$C = \frac{1}{T_{\text{RTS}} + \text{SIFS} + T_{\text{CTS}} + \text{SIFS} + T_{\text{DATA}} + \text{SIFS} + T_{\text{BACK}} + \text{DIFS} + T_e}, \tag{1}$$

All the parameters are collected in manuscript's Table A.3, located at the *Appendix A. Wireless Environment*. For clarification purposes, below we show the parameters used for data transmissions (Table 1).

Comment R2.3

The authors should compare MABs to other methods such as PID controller to maximize the reward function.

We appreciate the comment of the reviewer since it proposes another interesting approach for modeling the problem.

A PID controller is mostly used in industry to solve linear problems by continuously calculating an error function and providing actions based on predictions. In particular, a PID

Parameter	Description	Value
T_s	Symbol duration	9 μs
DIFS/SIFS	DIFS and SIFS duration	34 μs / 16 μs
N_{agg}	Number of packets aggregated	64
L_{DATA}	Length of a data packet	12000 bits
L_{RTS} / L_{CTS}	Length RTS and CTS packets	160 bits / 112 bits
L_{MAC}	Length MAC header	272 bits
L_{SF}	Length Service Field (SF)	16 bits
L_{MPDU}	MPDU delimiter	32 bits
L_{Tail}	Length tail	6 bits
L_{BACK}	Length block ACK	240 bits
T_{RTS}	RTS packet duration	$20 \cdot 10^{-6} + \frac{L_{\text{SF}}+L_{\text{RTS}}+L_{\text{Tail}}}{R} T_s$ s
T_{CTS}	CTS packet duration	$20 \cdot 10^{-6} + \frac{L_{\text{SF}}+L_{\text{CTS}}+L_{\text{Tail}}}{R} T_s$ s
T_{DATA}	Data packet duration	$36 \cdot 10^{-6} + \text{SUSS} \cdot 16 \cdot 10^{-6} + \frac{(L_{\text{SF}}+N_{agg} \cdot (304+L_{\text{DATA}})+L_{Tail})}{R} T_s$ s
T_{BACK}	Block ACK duration	$20 \cdot 10^{-6} + \frac{L_{\text{SF}}+L_{\text{BACK}}+L_{\text{Tail}}}{R} T_s$ s

Table 1: Simulation parameters

controller relies on three components, which are aimed to control the generated error. The output of the PID controller is generated as follows:

$$u(t) = K_p P + K_i I + K_d D = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt}, \tag{2}$$

where K_p, K_i and K_d are the coefficients provided to each component.

With regards to the decentralized SR problem presented in this work, we have thought about how it could be modelled by means of a PID controller:

- The experienced throughput is the Process Variable (PV).
- The desired throughput is the Set Point (SP), which in our work is referred to as optimal reward.
- Therefore, the error function refers to the regret, which is the difference between SP and PV.
- Finally, the selected action in terms of frequency channel, transmit power and sensitivity threshold is the output $u(t)$ from the PID controller in time t.

However, as shown by definition, the desired output for our system refers to a non-linear combination of independent parameters that impact on the error function. Moreover, the last term of the PID controller, meant for predicting the error, may not be used in our case, since the error function depends on an unknown environment.

We do believe, as highlighted by the reviewer, that comparison with other methods is important. So, we have provided additional results to Subsection *5.1. Selfish vs Environment-Aware Learning*, where the ε-greedy method has been executed for the same scenarios. We hope that this improvement helps at justifying the election of Thompson sampling instead of other learning-based methods. We also refer the reader to **?**, where we compared several methods for the decentralized SR problem in WLANs.

Comment R2.4

MIMO techniques such as adaptable beamforming and beam tracking (to increase coverage and minimize interference) in the range of actions (A) and their impact should be included.

Response. We strongly agree that techniques such as MIMO and beamforming would significantly contribute to the interference minimization problem. Our framework can be extended to incorporate these techniques, however substantial work is needed. Accordingly, we have modified Section *6. Conclusions* to provide a description on how our work could be extended to account for this.

In addition, we would like to introduce here a short description on the matter. On the one hand, we think that we could apply the same mechanisms as presented in the paper if we could assume that both MIMO and beamforming allow to relax the problem in a similar way than using several non-overlapping frequency channels. On the other hand, when referring to dynamic scenarios in which there are more STAs per WLAN, the application of MIMO and beamforming is expected to have stronger implications on the SR operation. In particular, much more complex interactions between devices may generate situations that require a more detailed examination. For instance, in case of applying beamforming, a different interference model would be generated than for the case of omnidirectional transmissions. Therefore, the number of overlapping WLANs may vary, thus impacting on their interactions. As a result, tuning both the transmit power and the sensitivity may have significantly different implications than for the default case. It is crucial, then, to properly understand these new interactions before providing a decentralized SR learning-based solution.

Continuing with the beamforming issue, we envision a potential SR solution on a per beam basis, rather than on a per network basis. The fact of using different beams for each associated STA motivates such an assumption.

Comment R2.5

Mobility of STAs should be included.

Response. We completely agree with this, and, since both Reviewers 1 and 2 addressed the same issue, we highly increase our interest to the matter. We want to refer to the Response granted to Comment R1.1 to address Comment R2.5.

We would like to point out that it is interesting to study whether the provided learning mechanisms are fast enough to maximize the performance of WLANs during *i)* stationary periods (i.e., when the network remains relatively stable), and *ii)* the transitory regimes (i.e., after a change in the environment occurred). Of course, such an analysis strongly depends on the average time that networks remain static. It is worth to mention that domestic Wi-Fi networks (which are mostly targeted in this work), remain stable for large periods.