Response Letter of *Collaborative Spatial Reuse in Wireless Networks via Selfish Multi-Armed Bandits*

Journal on Ad-hoc Networks

Francesc Wilhelmi, Cristina Cano, Gergely Neu, Boris Bellalta, Anders Jonsson & Sergio Barrachina-Muñoz
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 ADHOC_2018_155. First of all, 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 with the aim of solving 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. Before stating 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 restructured several parts of the article, thus providing a more logical flow, so that readability is enhanced.
- We have now broaden the related work with the purpose of granting a more general overview of the Spatial Reuse (SR) paradigm. Accordingly, we have provided a top-down approach to end up in the Reinforcement Learning (RL) potential solution.
- We have highlighted the main contributions of the article, and have reduced redundancy and shortened the parts that were too detailed.
- We have now further analyzed the variability issue arised from the decentralized learning operation. In particular, we have introduced two different learning procedures, *concurrent* and *sequential*, which operation is compared. In particular, learning sequentially contributes to significantly reduce the experienced temporal throughput variability of the concurrent scheme, even if the reward is still selfish.
- We have improved the evaluation part by *i)* concentrating results in fewer figures, *ii)* using more realistic parameters, *iii)* studying dynamic scenarios, and *iv)* including new results regarding the abovementioned proposed mechanisms.

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, Cristina Cano, Gergely Neu,
Boris Bellalta, Anders Jonsson & Sergio Barrachina-Muñoz
Barcelona (Spain), November 13, 2018

1 Reviewer #1

Comment R1.1

The work from the beginning addresses the spatial reuse issue with a much focused approach to machine learning. However, in Wireless Fidelity (WiFi) networks, channel selection and transmit power control are mechanisms considered since 2003 in the IEEE 802.11h extension. Moreover, still concerning spatial reuse, several approaches different from machine learning have been proposed (see, for example: (i) "CACAO: Distributed client-assisted channel assignment optimization for uncoordinated WLANs," IEEE Trans. Parallel Distrib. Syst., 2011, which relies on the exploitation of feedback information; (ii) "On the design of MAC protocols for multi-packet communication in IEEE 802.11 heterogeneous networks using adaptive antenna arrays," IEEE Trans. Mobile Comput., 2015, which relies on spatial filtering, and (iii) "Collision recovery in distributed wireless networks with opportunistic cooperation," IEEE Commun. Lett., 2010, which relies on cooperation). Furthermore, the entire literature involving multiple antenna systems is not considered, even if it has lead to the currently implemented 802.11n, 802.11ac, and 802.11ad extensions. In summary, the authors should arrive to the adopted machine learning approach by moving from a more general view of the spatial reuse issue, thus considerably extending the literature overview and, furthermore, clearly explaining the advantages of machine learning with respect to the other possible approaches (spatial filtering, spatial multiplexing, ...).

Response. We really appreciate the literature provided by the reviewer, which has contributed to enrich the motivation of the approach presented in this article. In particular, we modified Sections *1. Introduction* and *2. Related Work* to provide a broader view of Spatial Reuse (SR) solutions. Therefore, a more natural flow is now followed before reaching machine learning as a potential solution.

In particular, we find that the decentralized SR problem presented in this work, which is based on Transmit Power Control (TPC) and Dynamic Channel Allocation (DCA), leads to many complex interactions among Wireless Networks (WN). Those interactions are very hard to model in practice, since they can be found on the spatial domain and are mostly generated by additive interference. Moreover, the decentralized setting that we focus on prevents to provide "full information"-based solutions like the one that is shown in Tang et al. (2014). All of this makes us think in Reinforcement Learning (RL) as a potential solution to the abovementioned problem, which feasibility is aimed to be shown in this paper. RL allows reducing the complexity of the problem by getting rid of specific information of every wireless node in a given scenario. In contrast, we propose to decentralize the problem by allowing each WN to gain knowledge on all the other adversarial networks as a whole, thus facing a single environment. Accordingly, a higher flexibility can be obtained with respect to the scenario: learning is done according to the current situation of nodes, which not only involves potential interactions but environmental conditions (e.g., propagation effects). Of course, the application of decentralized RL entails some benefits and drawbacks, which this paper aims to shed light on.

Comment R1.2

A second key point concerns the research contribution. This work substantially represents an interesting overview of existing MAB methods, which are implemented to check their performance. Thus, it is mainly of tutorial nature. The partial novelty lies in the application of these methods to the wireless network context, but, in the referee's opinion, this is not sufficient for a research paper. Within the proposed study, a novel method, an extension of an existing method, or a combination of methods should be proposed to improve the scientific contribution of the manuscript. In this sense, the investigation of the consequences due to the adversarial setting in the presence of a not planned network deployment, claimed in the first paragraph of Section 3, should be also better highlighted in the Introduction and in the Abstract.

Response. Following the reviewer's comment, we have provided an enhancement of the current decentralized solutions to alleviate the problem of the temporal throughput variability. Recall that the throughput variability stands for the fluctuations in the throughput experienced by a given WN, which are entailed to negatively impact both the user's experience and the upper communication layers.

In particular, in order to study the variability generated when learning, we have considered two learning approaches: concurrent (as in the previous version of the paper) and sequential (added in this version). By doing so, we aim to shed light on the kind of mechanisms that would contribute to improve the fully decentralized situation. In particular, Section *4.3 Learning Procedure* describes in detail both the concurrent and the sequential learning mechanisms, which operate on the top of the learning strategies presented in this work.

We also would like to clarify in this response letter the motivation behind presenting a comparison between concurrent and sequential learning approaches. The fact is that we initially attempted to frame the fully decentralized problem, i.e., without any type of coordination, which would correspond to the concurrent approach. However, some performance issues were noticed, especially in terms of temporal throughput variability. To mitigate this, we have analyzed the effect of taking actions concurrently and sequentially. Figure 1 in the document illustrates both mechanisms, which are now evaluated in Section *5. Performance Evaluation*.

Finally, we have modified the article (with especial emphasis on the *Introduction* and the *Abstract*) to highlight the actual contribution of this work. The main contribution, as Reviewer #1 properly says, mostly refers to the analysis of the implications of applying online learning in unplanned deployments.

Comment R1.3

The paper seems too long for the content it provides, with many details regarding each action-selection strategy that reduces the readability of the work. The description should be more concise and focused on the most relevant mechanisms of each strategy. Besides, the presentation of the results, with a huge number of very small subfigures, makes difficult to infer the most relevant differences among the various strategies. A reduced number of clarifying tables/figures may be highly preferable.

Response. We find this comment very valuable since it allowed us to take a more critical view on what was written. Accordingly, we have applied significant changes to the article. In particular, the following Sections were reworked:

- Section *3.2. Mult-Armed Bandits Formulation for Decentralized Spatial Reuse* has been considerably reduced and restructured together with Section *4. System Model*. Particular emphasis has been given to the first part of Section *3.2.*, where the decentralized SR problem is modeled through MABs. However, we decided not to remove the implementation details of each algorithm (corresponding to subsections *3.2.1.*, *3.2.2.*, *3.2.3.* and *3.2.4.*, respectively), since we find this "mini-tutorial" part to be an interesting contribution.
- Section *4. System Model* has been reworked to provide only the essential information. Note that the details regarding the learning procedure have been moved from Section *3. Multi-Armed Bandits for Improving Spatial Reuse in WNs* to subsection *4.3. Learning procedure*, which we consider is better for the readability of the article.
- Significant structural changes have been provided to Section *5. Performance Evaluation*.
  - First, subsection *5.1. Toy Grid Scenario* has been shortened and the provided results have been concentrated to a smaller number of figures. In particular, we removed the subsections dedicated to each algorithm. We have divided this subsection in two: one referring to the parameters tuning, and another one for comparing the algorithms' performance.
  - Second, subsection *5.2. Random Scenarios* has been adapted to include the new results (refer to comments R1.2, R1.4, R2.1 and R2.3). In addition, results are now more compact and figures aim to be more meaningful.

Comment R1.4

A final relevant discussion should be inserted regarding the usefulness of machine learning in providing spatial reuse with respect to beamforming or cooperation, specifically considering the existing papers (i.e., (i)-(iii) and many others), and also the existing standard extension (i.e., 802.11h/n/ac/ad). This latter discussion should be supported by numerical results at the end of Section 5.

Response. We strongly agree with this comment and think that the ML operation should be compared with other techniques. In particular, beamforming and/or MU-MIMO are very powerful techniques that may contribute to SR enhancement. However, its application requires from an exhaustive analysis that we decided to keep out of the scope of this work. Nevertheless, we have considered the comparison of our spatial reuse solution with the current IEEE 802.11 operation. More specifically, we have added "static" results to Figure 8 (subsection *5.2. Random Scenarios*). By static we refer to the situation where parameters are not modified online, thus framing typical residential and chaotic deployments, where the usage of resources is sub-optimal.

Finally, we would like to add some additional comments regarding the possible inclusion of both MU-MIMO and beamforming to the SR problem presented in this work. First of all, we think that we could apply the same mechanisms as presented in the paper if we could assume that both MIMO and beamforming allow relaxing the problem in a similar way than using several non-overlapping frequency channels. However, its application is expected to have important implications with regards to the interactions among WNs.

In particular, adding MIMO and/or beamforming to the SR problem may become significantly challenging for scenarios with multiple STAs. In such a situation, different solutions should be provided for each pair of transmitter-receiver pairs. For instance, in case of applying beamforming, the interference model would differ from the case of omnidirectional transmissions (which is the one that is studied in this paper). As a result, the interference sensed at a given node would vary according to the transmitter-receiver pairs that are currently transmitting. In practice, tuning the transmit power uniformly (i.e., the APs use the same configuration for all its associated STAs) would generate these multiple interactions among nodes. Accordingly, applying the RL methods as proposed in this work may lead to unpredictable behaviors. 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.

To conclude, it is crucial to properly understand the new interactions generated by multiple antenna strategies before providing a decentralized SR learning-based solution.

2 Reviewer #2

Comment R2.1

One of the reason for formulating the resource allocation problem as a MAB is the absence of coordination among the WNs for solving the problem. However, it would be worth evaluating distributed techniques for solving the resource allocation problem based on message passing algorithms like gradient descent and primal-dual interior-point methods. Amendments like 802.11f already provide protocols for the communication among APs. Therefore the implementation of a message passing mechanism is practicable. I would suggest authors to compare their solutions with an iterative algorithm for distributed optimization based on message passing (using for example primal-dual methods).

Response. We really appreciate this comment, which, in addition, has been similarly posed by Reviewer #1. For that reason, we paid especial attention to this matter and added significant changes to improve the article's contribution. In particular, we provided two different learning approaches, namely *concurrent* and *sequential* learning, which operate on top of the proposed algorithms. Sequential learning, as suggested in Comment R2.1, is based on message passing, and is aimed at reducing the temporal variability that is shown to be unleashed in the adversarial setting. Note, as well, that we preserved the nature of the decentralized SR problem by using the selfish reward in both concurrent and sequential approaches. Finally, we would like to remark that collaborative mechanisms have been considered as part of the future work. In fact, a collaborative message-passing approach has been already presented in Wilhelmi et al. (2018).

The definition of concurrent and sequential learning has been provided into a new Section

(*4.3. Learning procedure*). In addition, the results in Section *5. Performance Evaluation* have been complemented accordingly.

Comment R2.2

The proposed channel selection/switching policy based on MAB algorithms seems to take too long to converge. If we assume that the AP change the channel once every beacon interval, by using the beacon to inform the stations about the channel switching, it would takes approximately 16 minutes before convergence (a beacon is transmitted every 100ms, which results in 1000s for 10k iterations). Is there any way to speed up the convergence? Which is the performance loss if we stop the execution of the MAB procedure in before convergence?

Response. Regarding the convergence issue, we agree that 16 minutes seems to be too long. On the one hand, we refer to adversarial settings, so convergence is not guaranteed. However, as shown in the paper, WNs reach some sort of equilibrium given certain action-selection strategies. In addition, learning sequentially contributes to achieve such a convergence. On the other hand, we assume that a potential use case of the kind of mechanisms presented in this work could be residential (and independent) wireless LANs, which are characterized by being active for long periods. As a result, the proposed decentralized algorithms are expected to improve their performance in the long term.

When it comes to the possibility of speeding up convergence, ongoing work is currently in progress. In particular, the convergence analysis and its implications on the performance of wireless networks is further studied in Wilhelmi et al. (2018).

Finally, regarding the last of the questions addressed by Reviewer #2 in this comment, significant changes have been added to subsection *5.2. Random Scenarios*. The fact is that we now aim to show the gains achieved by applying MABs for different intervals of iterations. To that purpose, Figure 8 has been included to illustrate the average throughput achieved in the random scenarios. By showing the gains achieved during each interval, we expect to provide insights on the performance gains of each strategy, even though convergence may not be guaranteed (due to the scarcity of the resources).

Comment R2.3

The analysis is mainly limited to a toy example that permits to get to the gist of the analyzed techniques, but it does not provide any evidence that these solutions are suitable for realistic networks. The convergence speed (i.e., the rate at which the regret fades away) depends on the size of the instance. A larger and more interconnected decision space requires more exploration, thus reducing the scalability of the analyzed techniques. I wonder how these techniques perform in more realistic networks. Even if co-existing WiFi networks are rather limited, the number of channels and power levels is higher than those considered in this work. I would suggest authors to consider larger instances by increasing the number of channels, power levels and connected stations.

Response. According to the reviewer's comment, we tried to provide more realistic simulations through the following parameters:

- Number of orthogonal channels: 3, corresponding to the 2.4 GHz band.
- Transmission power levels (in dBm): {-15, 0, 15, 30}.

All the simulations presented in Section *5. Performance Evaluation* have been re-done according to the new parameters, thus leading to different results. Note that we provided one new orthogonal channel, which significantly increased the action space. In particular, for the toy grid scenario (4 WNs), we have now 20,736 possible combinations of actions (joint profile), in contrast of the 4,096 combinations that we had at the time of submitting this article for the first time. It is worth noting that increasing the number of orthogonal channels helps with convergence speed, since more feasible solutions become available.

In contrast, we decided not to increase the granularity of the transmit power actions with the aim of illustrating the actions probabilities in Figure 5 (subsection *5.1.2. Performance of the MAB-based Policies*). However, we provided values compliant with the IEEE 802.11 amendment (Kasslin and Lappeteläinen (2000)).

Finally, we would like to mention that we attempted to make more emphasis on the results provided in Section *5.2. Random Scenarios*, where random and dense scenarios are provided. Up to 8 overlapping WNs are considered to coexist within the same area as for the toy scenario.

Comment R2.4

MAB algorithms have the ability to adapt to a changing environment. However, the convergence speed after a change highly depends on the ability of the algorithm to perform the exploration phase. However, the empirical analysis is limited only to a static scenario. I would suggest authors to consider also a dynamic scenario to see how these algorithms react to a change in the distribution of STAs and interference.

Response. Motivated by this comment, we have included additional content regarding a dynamic environment, which is now found at Section *5.1.4. Learning in a Dynamic Environment*. What we did is to provide a dynamic scenario in order to analyze the performance of the best-performing learning strategy, i.e., Thompson sampling. The dynamic scenario is based on the toy scenario, but WNs are gradually turned on. With that, we aim to show the ability of the MAB-based strategies to adapt to changes in the environment (i.e., nodes arrivals).

In addition, we would like to remark that applying Reinforcement Learning to dynamic environments is a challenging task that we are currently exploring. In particular, we find variations of the algorithms presented in this work that are meant for changing environments. That is the case of Dynamic Thompson Sampling (DTS) (Gupta et al. (2011)), which has been shown to adapt faster to changes in the environment than the traditional Thompson sampling. Roughly, DTS promotes adaptive exploration by tracking the reward probabilities of each arm, which is useful to give emphasis to recent observations.

References

Gupta, N., Granmo, O.-C., and Agrawala, A. (2011). Thompson sampling for dynamic multi-armed bandits. In *2011 10th International Conference on Machine Learning and Applications Workshops*, pages 484–489. IEEE.

Kasslin, M. and Lappeteläinen, A. (July 2000). Transmitter power control (tpc) for 802.11 wlan - rev.1. *doc.: IEEE802.11-00/190*, pages 802–11.

Tang, S., Yomo, H., Hasegawa, A., Shibata, T., and Ohashi, M. (2014). Joint transmit power control and rate adaptation for wireless LANs. *Wireless personal communications*, 74(2):469–486.

Wilhelmi, F., Barrachina-Muñoz, S., Cano, C., Bellalta, B., Jonsson, A., and Neu, G. (2018). Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in wlans. *arXiv preprint arXiv:1805.11083*.