Learning based Spatial Reuse with Adaptive Timestep and Action Space for Dense WLANs

ZHAO WEN CHOW[1] SHOTA SAKAI[1] HIROSHI SHIGENO[1]

Abstract:
The rapid densification of IEEE 802.11 Wireless Local Area Networks (WLANs) has lead to higher interferences among Basic Service Sets (BSSs) and has negatively impacted their performance. Spatial reuse methods such as Dynamic Sensitivity Control (DSC) or Transmit Power Control (TPC) help mitigate the hidden and exposed terminals issues in these dense deployments. In this work, a Reinforcement Learning (RL) based method with adaptive timestep and action space is proposed to enhance the spatial reuse in dense WLANs. In particular, the problem is modeled through Multi-Armed Bandits (MABs) and the Thompson Sampling strategy is employed. In this scheme, a learner first observes the Received Signal Strengths (RSSs) it can sense and derives a set of Carrier Sense Thresholds (CSTs) from these. It then applies Thompson Sampling with the computed set and updates the model after a specified number of transmissions or a predefined timeout. Simulation results show that the proposed scheme is able to improve the fairness compared to a previous RL scheme while providing a considerable aggregate throughput.

1. Introduction

In recent years, the increasing demand for wireless communication encouraged more WLANs deployments and as a result, the overall performance degraded due to higher interferences among basic service sets (BSSs). In particular, the popular IEEE 802.11 standard, which implements Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), experiences severe throughput degradation and unfairness in dense deployments [1]-[3].

Two issues arising in dense scenarios are the hidden and exposed terminal problems [4]. The hidden terminal problem consists in a collision between two packets at the receiver due to the transmitters not being able to sense each other and transmitting simultaneously. On the other hand, an exposed terminal is a node which experiences low transmit opportunities due to excessive carrier sensing.

Two spatial reuse methods for improving the performance under dense deployments are Transmit Power Control (TPC) and Dynamic Sensitivity Control (DSC). TPC consists in adapting the transmit power while DSC adapts the Carrier Sense Threshold (CST). Both methods aim to optimize the spatial reuse of radio resources and decrease hidden and exposed terminals for improved performance.

However, tackling the spatial reuse problem with TPC and DSC is not a trivial task, especially in uncoordinated environments. As the density of WLANs increase, the interactions between uncoordinated nodes become very complex to model and it is therefore very challenging to devise a performing algorithm. Moreover, wireless communications are dynamic by nature which adds another layer of complexity. Although many previous studies use traditional algorithms, there has been an increasing interest in learning based methods for applying spatial reuse [5]-[7]. In fact, Machine Learning (ML) has been a very promising technology over the last years as it has proved its potential to achieve results similar or even better than classic approaches without requiring overly complex models and much knowledge of the environment. In particular, Reinforcement Learning (RL) is a method where the learning is carried on while interacting with the environment and the computational cost can be kept low compared to other learning algorithms. Learning based methods are therefore appealing for dynamic environments such as WLANs.

In this paper, we continue on studying the potential of using RL for improving the throughput and fairness in dense WLANs scenarios. In particular, we focus on IEEE 802.11 networks in uncoordinated environments, e.g. residential buildings. We propose a spatial reuse scheme based on Multi-Armed Bandits (MABs) which adapts its learning timestep and action space to efficiently control the CST at each Access Point (AP). First, each learner, i.e. AP, observes all the Received Signal Strengths (RSSs) it is able to sense. It then derives a set of CST values from these observed RSS values and performs the Thompson Sampling algorithm to find the optimal CST from this set. Every learning step consists in selecting one CST and observing the experienced throughput when applying this threshold for a certain number of transmissions or until a predefined timeout. Adapting the timestep and the set of CSTs (actions) for each learner allows to focus on relevant threshold values and improve the learning process.

[1] Graduate School of Science and Technology, Keio University, Yokohama, Kanagawa, 223-8522, Japan

The remainder of this paper is structured as follows: related works are presented in Section 2. The MAB framework is detailed in Section 3. The proposed scheme is specified in Section 4. The performance of the proposed scheme is evaluated in Section 5. Finally, the conclusion of this paper is given in Section 6.

2. Related Work

In the literature, researchers have applied TPC, DSC or both to enhance spatial reuse in dense scenarios [8]-[10]. Although many algorithms are traditional, in the sense that they do not use ML, there has been an increase in learning based methods [5]-[7].

I. Jamil et al. [5] proposed a centralized solution based on multilayer perceptron to jointly adapt the transmit power and the CST. Their artificial neural network architecture aims to model the relationship between the throughput achieved by the nodes and their transmit power and CST. It is composed of the input layer, 1 hidden layer and the output layer. The cost function is the sum of a fairness cost and a throughput cost as they aim to achieve a minimum average throughput per device while maintaining fairness. Although their solution showed an improvement in the aggregate throughput and the fairness, it requires a central controller to perform the learning. This is not necessarily feasible in uncoordinated environments such as residential scenarios.

F. Wilhelmi et al. [6] proposed a decentralized Reinforcement Learning approach to spatial reuse. More specifically, the authors apply a stateless variation of Q-learning to control the transmit power and the channel used based on the experienced throughput. Their approach showed the potential of improving the aggregate throughput although the individual throughputs experience high variability due to the competition among learners. The algorithm's performance was evaluated in a relatively simple scenario, containing few nodes, mainly to check the potential of using RL for spatial reuse so further study is necessary.

The work of F. Wilhelmi et al. [7] considers an RL approach based on Multi-Armed Bandits (MABs) and the Thompson Sampling action-selection strategy. In their framework, each learner has the objective of maximizing its throughput by exploring and finding the optimal combination of transmit power, CST and frequency channel. They evaluated two selection strategies: selfish and environmental-aware learning. The first one consists in having a reward based solely on the throughput experienced by the node itself. In the second case, the reward is based on the max-min throughput achieved in the entire network. They showed that selfish learning has the potential of maximizing the aggregate performance although it can generate unfair situations. On the other hand, environmental-aware learning allows to solve fairness issues but does not guarantee the optimal solution and may drastically limit the aggregate performance. It is also worth noting that their environmental-aware approach assumed perfect estimation of the neighboring nodes' throughput, which may not be feasible or accurate in real settings. Moreover, they do not specify any timestep and assume a perfect estimation of the expected rewards and their simulations only use a simple set of 2 CSTs for all learners.

In summary, although learning based spatial reuse methods of previous studies could achieve improved performance, it remains a challenge to achieve high aggregate throughput while preserving fairness among nodes in uncoordinated environments. Selfish decentralized learning is prone to generate unfair situations while collaborative learning seems promising to maintain fairness but needs to overcome some practical issues.

3. Multi-Armed Bandits

The MAB problem is a classic RL problem where resources need to be allocated to a set of choices in a way to maximize the reward in the long run. The learner only has limited knowledge about each choice at the time of allocation and acquires additional information about one choice the more it allocates resources to it. In the context of spatial reuse, a node trying to learn which parameters, such as the transmit power or the CST, maximizes its throughput can therefore be described by the MAB problem. In the case where there are multiple learners and they compete for the same resources at the same time, it can be modeled as an adversarial MABs. This is exactly the situation where multiple Wireless Local Area Networks (WLANs) contend for accessing the channel and the transmit power and CST are two parameters which influence this contention.

It has been shown that the MAB framework is able to deal with the exploration-exploitation dilemma under high uncertainty [11]. It is therefore suitable for decentralized spatial reuse as each node has very little or no information regarding the environment. Moreover, it is worth noting that wireless communications are sensitive to delay and more computationally intensive algorithms such as Deep Learning are thus not feasible.

3.1 Thompson Sampling

One existing selection strategy which has proved to efficiently address the exploration-exploitation trade-off in the context of wireless networks is Thompson Sampling (TS) [12]. It is a Bayesian algorithm which builds a probabilistic model of the expected reward for each action. Initially, it assumes a prior distribution for the expected reward of each action and after playing a certain action, the model is updated with the posterior distribution given the actual observed reward. For each learning step, every action has a probability of being selected which match its probability of being the optimal action. This strategy can be implemented by sampling from the posterior distribution of each action and playing the one associated with the sampled value yielding the highest expected reward.

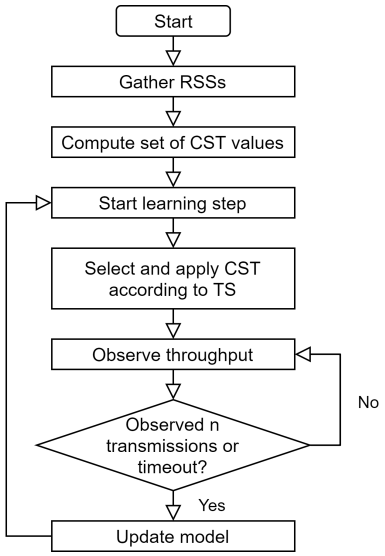

Fig. 1 Algorithm of the proposed scheme

Algorithm 1 Thompson Sampling

Input: Set of possible actions $\{1, ..., K\}$

Initialize:
  $t \leftarrow 0$
  $\hat{r}_k(t) \leftarrow 0,\ k = 1, ..., K$
  $n_k(t) \leftarrow 0,\ k = 1, ..., K$
while true do
  Sample $\mathcal{N}\left(\hat{r}_k(t), \sigma_k^2(t)\right)$ and get $\theta_k(t),\ k = 1, ..., K$
  Play action $k = \underset{k}{\operatorname{argmax}}\, \theta_k(t)$
  Observe the throughput
  Compute the reward $r_k(t)$
  $\hat{r}_k(t) \leftarrow \dfrac{\hat{r}_k(t) \cdot n_k(t) + r_k(t)}{n_k(t) + 1}$
  $n_k(t) \leftarrow n_k(t) + 1$
  $t \leftarrow t + 1$
end while

4. Learning based Spatial Reuse with Adaptive Timestep and Action Space

In this section, we propose a scheme based on Thompson Sampling which adapts its learning timestep and action space to allow an agent to efficiently learn its optimal CST and to enhance the spatial reuse. An overview of the proposed scheme is shown in Figure 1. The Thompson Sampling algorithm is also given in Algorithm 1.

In this work, the agents are the Access Points (APs). Each AP has the goal to learn its optimal CST from a set of CST choices to maximize a certain reward function based on the experienced throughput. This set of CST choices depends on the perceived RSSs of each learner as it will be explained later. Regarding the throughput, it is the throughput in transmission and only packets which were successfully transmitted, i.e. for which an acknowledgement (ACK) has been received, are counted in it.

As mentioned in Section 3, Thompson Sampling starts by assuming a prior distribution for the expected rewards and in this scheme, a Gaussian prior is assumed, similar to the work of Wilhelmi et al. [7]. The posterior distribution in this case is thus also Gaussian with mean $\hat{r}_k(t)$ and variance $\sigma_k^2(t)$

$$\hat{r}_k(t) = \frac{\sum_{w=1}^{t} r_k(w)}{n_k(t) + 1} \quad \text{and} \quad \sigma_k^2(t) = \frac{1}{n_k(t) + 1}, \tag{1}$$

where $n_k(t)$ is the number of times action k was played until time t.

Regarding the reward, the selfish reward presented in [7] is reused in this work. The latter defines the reward $r_w(t)$ of a learner w at timestep t as

$$r_w(t) = \frac{\Gamma_w(t)}{\Gamma_w^*} \tag{2}$$

where $\Gamma_w(t)$ is the throughput and Γ_w^* is an upper bound value for normalization. This upper bound may not be known by the learner as it depends on the spatial configuration of the nodes as well as many other parameters. Consequently, the maximum throughput achievable, i.e. when there are no interferences, given the Modulation and Coding Scheme (MCS) is used as the upper bound instead.

4.1 Adaptive timestep

The duration of one learning step is a crucial parameter in the Thompson Sampling algorithm and should be wisely chosen as it directly influences the learning process. In fact, a short timestep avoids playing bad performing action for a long period of time and allows more exploration but provides a less accurate long-term estimation of the action's performance. A long timestep yields the opposite.

One simple way to implement Thompson Sampling is to try different fixed timesteps and select the best. However, this method could yield inaccurate performance estimations. For example, if a learner starts to send a packet during a learning step but does not finish before the end of the step, this packet will not be counted in the throughput computation of the current step even if it succeeds later on. A long timestep mitigates this issue as the number of packets observed is increased so that the last one should have minimal influence on the throughput but as noted above, the amount of exploration is decreased so that the performance estimations are less accurate. Moreover, the "best" timestep could vary from scenario to scenario. A more complex situation with many WLANs should require a longer duration to acquire a good long-term estimate than a more simple one.

In this work, a variable timestep has been considered. The idea is to observe n transmission(s) or until a predefined maximum timeout t_o during one learning step. The latter is therefore adjusted to some degree to a learner's local environment. One transmission ends when either the learner receives an ACK or timeouts if the ACK is not received. Moreover, the additional timeout t_o prevents a learner getting stuck with one action indefinitely.

4.2 Adaptive action space

The set of CST choices is another aspect which influences the learning process. Essentially, the Thompson Sampling

algorithm considers each action as independent from the others and having its own reward distribution. In consequence, fewer actions decreases the amount of exploration and leads to more accurate estimations although more performing actions might exist. On the other hand, more actions could potentially lead to the optimal performance but involves more time in exploring all the different possibilities. To illustrate this slowdown, suppose there is a fixed amount of learning steps n. If there is only one action, every step will play this value and the variance at the end will be $\sigma^2 = \frac{1}{n+1}$ (Eq. 1). If there are two actions, one action will be selected at maximum n times but will be in general less than n. Its variance is thus higher meaning the estimation of the expected reward is less accurate.

Some choices might also be similar in terms of performance. For example, a CST of -80 dBm or -79 dBm will yield the same result if a node does not sense any signal between these two values. Furthermore, every learner experiences different RSSs depending on the location of their neighbors so a distinct set of CST choices for each one could potentially lead to more efficient learning.

Real life wireless communication is subject to fading and shadowing so the perceived RSSs are constantly changing in time. However, the study of these effects is out of the scope of this paper. Therefore, they are assumed to be constant hereafter.

Based on the ideas above, the proposed scheme implements a different set of CST possibilities for each learner which depends on their perceived RSSs. The process is divided into two phases: an initial phase and a learning phase. During the initial phase, each learner does not apply RL and saves all the different RSSs [dBm] it senses. This set of saved values S (different for each AP) is then used to derive the set of CST choices A as follows:

$$A = \{g(s) \,|\, s \in S \text{ and } -82 \leq s\} \tag{3}$$

with $g\colon [-82, \infty) \to [-82, -62]$ defined as

$$g(x) = \begin{cases} \lfloor x \rfloor, & \text{if } x < -62 \\ -62, & \text{otherwise} \end{cases} \tag{4}$$

The number of perceived RSSs, i.e. sensed neighbors, is thus different for each CST choice. The use of the floor function limits the number of choices (actions) to avoid a large action space which would require a lot of exploration.

Once the set of actions is derived, the learners enter the learning phase and apply Thompson Sampling with these actions.

5. Evaluation

5.1 Simulation settings

The proposed scheme was evaluated using the ns-3 simulator [13] in 50 random scenarios based on the residential scenario described by IEEE 802.11ax TG [14]. Every scenario consists in one floor of 10 × 2 apartments, each of size 10 m × 10 m × 3 m as shown in Figure 2. One BSS is placed randomly in the xy plan in each room and consists of one AP and one station (STA). All nodes are at height $z = 1.5$ m from the floor and have a static mobility. An overview of the simulation parameters are shown in Table 1. Regarding the number of transmissions considered for the learning step, simulations for various values have been performed to evaluate its influence and find an optimal value.

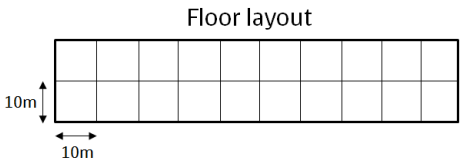

Fig. 2 Scenario layout

Table 1 Simulation parameters

Wi-Fi standard	802.11ac
Frequency band	5 GHz
Channel number	38
Channel bandwidth	20 MHz
Spatial stream(s)	1
MCS	7
Rate control	None
Propagation loss	Residential building loss [14]
Shadowing	None
Mobility model	Static
Traffic model	CBR
Traffic load	Full buffer DL
MPDU size	1544 bytes
Max Aggregation	64
RTS/CTS	Disabled
Max retransmissions	7
Transmit power	AP: 23 dBm, STA: 15 dBm
CST	AP: -82 dBm (initialization), STA: -82 dBm
Antenna gain	AP: 0 dBi, STA: 0 dBi
Capture effect threshold	5 dBm
Floor noise level	-101 dBm
Noise figure	7 dBm
Simulation duration	Initialization: 10 s, learning: 100 s

The performance metrics used are the aggregate throughput and the Jain's Fairness Index (JFI). The latter is computed as

$$\mathcal{J}(x_1, x_2, ..., x_n) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \cdot \sum_{i=1}^{n} x_i^2}, \tag{5}$$

where x_i is the throughput experienced by the ith AP and n is the total number of APs. Moreover, the proposed scheme is compared to the standard IEEE 802.11ac method (Legacy), which uses a static maximum transmit power (23 dBm) and carrier sensitivity range (-82 dBm), and the work of Wilhelmi et al. [7]. For the latter, the selfish reward, a fixed timestep of 0.5 s and an action set composed of CST values only $\{-82, -77, -72, -68, -62\}$ have been used.

5.2 Influence of the number of transmissions

Figures 3 and 4 show the mean and standard deviation of the aggregate throughput and the JFI obtained using the proposed scheme with a learning step of 1, 4, 20, 40 and 60 transmission(s). The throughput is maximal when con-

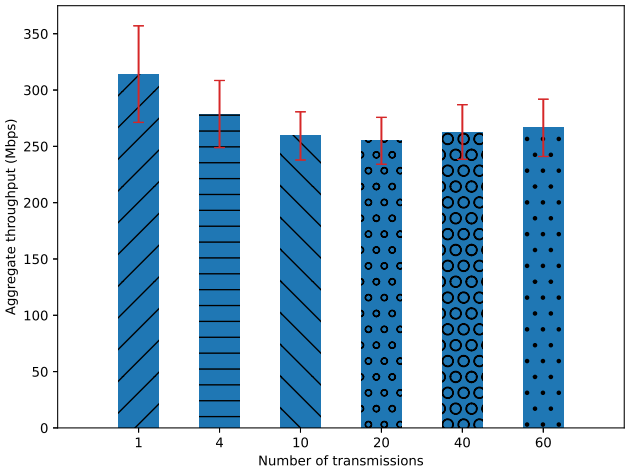

Fig. 3 Average aggregate throughput of the proposed scheme for different number of transmissions

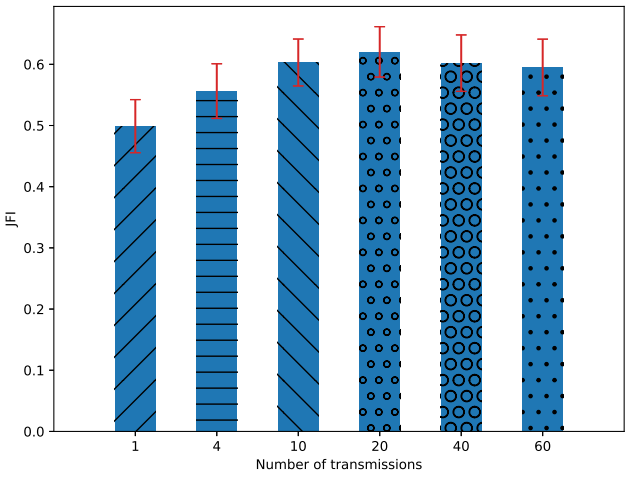

Fig. 4 Average Jain's Fairness Index of the proposed scheme for different number of transmissions

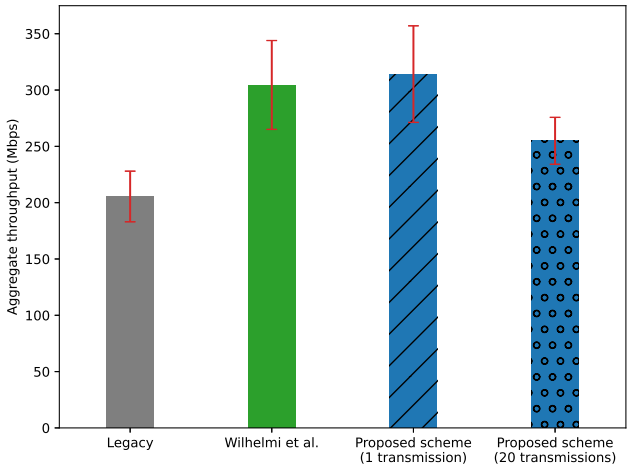

Fig. 5 Average aggregate throughput for all schemes

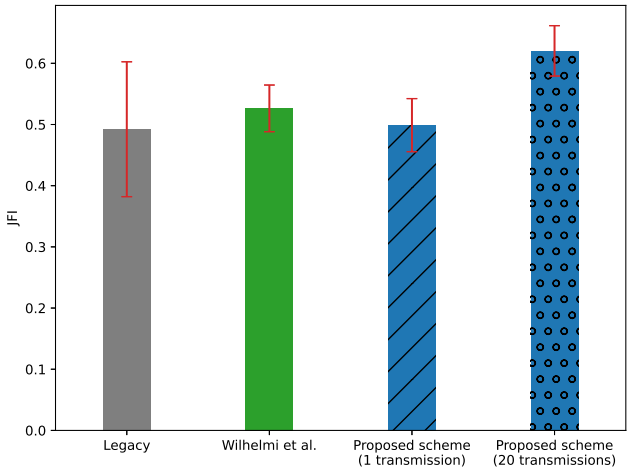

Fig. 6 Average Jain's Fairness Index for all schemes

sidering only 1 transmission and decreases as the number of transmissions considered increases until 20. It increases again afterwards.

Regarding the fairness, it increases when the number of transmissions observed increases until 20 and decreases again afterwards. Consequently, there seems to be a trade-off between the aggregate throughput and the fairness. One way to understand this is as follows: if a small number of nodes monopolize the channel and are able to experience maximum throughput for extended period of time, the total throughput will therefore be high at the detriment of some nodes which cannot communicate at all. Although some nodes will naturally experience lower throughput due to their poor location, it is essential to maintain a certain level of fairness.

Hereafter, only 1 and 20 transmissions will be retained as they yield the highest aggregate throughput and fairness respectively.

5.3 Comparison with previous schemes

The mean and standard deviation for the aggregate throughput and Jain's Fairness Index obtained after 50 random scenarios for each of the considered schemes are shown in Figures 5 and 6. First, the proposed scheme with 1 transmission yields a throughput improvement of 108.66 Mbps (52.88% increase) and 9.64 Mbps (3.17% increase) compared to the legacy and Wilhelmi's schemes respectively. In terms of fairness, it yields a similar fairness (only a 1.35% increase) as the legacy scheme while performing worse (by 5.21%) than Wilhelmi's scheme. When considering 20 transmissions, the throughput increased by 49.40 Mbps (24.04%) compared to the legacy algorithm while it is worse (by 16.29%) than Wilhelmi's scheme. Regarding the fairness, it is 26.01% and 17.86% higher compared to the same previous schemes. It is also worth mentioning that the standard deviation in the aggregate throughput is smaller for the proposed scheme with 20 transmissions than the other RL methods. In addition, the variability is similar for all learning schemes when looking at the fairness and smaller than the legacy case. The proposed scheme with 20 transmission performs thus more consistently across all random scenarios.

Wilhelmi's scheme apply Thompson Sampling with the same set of actions for all learners and a fixed timestep and is able to highly increase the aggregate throughput (48.18%)

compared to the legacy scheme. However, it does not improve much the fairness (6.92%). On the other hand, our proposed scheme which uses a different set of actions and a variable timestep yields a higher aggregate throughput but a lower fairness when observing for 1 transmission and vice-versa for 20 transmissions. In particular, their scheme always observes for the same duration (0.5 second) for every agent and action and the beginning of every step is done at the same moment for every learner (synchronized). In our proposed scheme, the beginning of each step can vary from learner to learner and the observation time is at most 0.5 second but can be shorter if a predefined number of transmissions have been observed.

In fact, across all random scenarios, our scheme observes for 0.0190 and 0.3009 second in average when considering 1 and 20 transmissions respectively. Despite having an average timestep much smaller than Wilhelmi's scheme, our proposal with 1 transmission performs roughly the same as theirs. One explanation is that a shorter timestep results in more learning steps performed during the same amount of time so the standard deviation for the expected rewards (Eq. 1) could be smaller. This potentially means a more accurate estimation of the actions' performance. However, a too small timestep would considerably lower the accuracy of the estimated long-term reward and seems to have a detrimental effect regarding the fairness.

Furthermore, our proposed scheme focus the training only on the relevant CST values for each learner instead of a common defined set used for Wilhelmi et al.'s work. Every AP in the simulation scenario is subject to a different spatial configuration of its neighbors so it will perceive different RSS values. As mentioned in Section 4, two CSTs for which the set of sensed neighbors is the same will likely produce similar performance. Our scheme prevents such happening and improve the learning process, thus leading to a higher performance.

In summary, the simulation results show that the basic IEEE 802.11 scheme does not exploit the spectrum resources to their fullest and a more efficient usage of these is possible via our learning based spatial reuse scheme. With our proposal, each AP efficiently learns its optimal CST by using an adaptive learning step and an appropriate set of CST values so that both the aggregate throughput and the fairness are improved compared to the legacy scheme. The proposed scheme also outperforms Wilhelmi et al.'s one [7] in terms of fairness by sacrificing some throughput while using a selfish reward.

6. Conclusion

In this paper, we proposed a Reinforcement Learning based spatial reuse scheme with adaptive timestep and action space to perform Dynamic Sensitivity Control. A learner first gathers all its perceived RSSs during an initialization phase then computes the set of CST values which will be used during the learning phase. During the latter, the Thompson Sampling algorithm is performed and each learning step lasts until a certain number of transmissions have been observed or a predefined timeout is reached. The proposed scheme has been evaluated through simulations and the results demonstrated the potential of applying Reinforcement Learning to improve the aggregate throughput while maintaining fairness in dense WLANs. In particular, an increase of 24.04% and 26.01% in the overall throughput and fairness respectively could be achieved compared to the legacy IEEE 802.11ac scheme. When compared to a previous RL scheme, it achieves a higher fairness by sacrificing some of the aggregate throughput.

Future work could consider the collaboration between learners in order to further improve the fairness among them. Moreover, the learning could be extended to take into account the configuration of the stations.

References

[1] N. Jindal et al., "Performance Gains from CCA Optimization," *doc.: IEEE 802.11-14/0889r3*, 2014.

[2] I. Jamil et al., "MAC simulation results for Dynamic sensitivity control (DSC - CCA adaptation) and transmit power control (TPC)," *doc.: IEEE 802.11-14/0523r0*, 2014

[3] M. S. Afaqui, E. Garcia-Villegas and E. Lopez-Aguilera, "IEEE 802.11ax: Challenges and Requirements for Future High Efficiency WiFi," in *IEEE Wireless Communications*, vol. 24, no. 3, pp. 130-137, June 2017.

[4] K. Nishide, H. Kubo, R. Shinkuma and T. Takahashi, "Detecting Hidden and Exposed Terminal Problems in Densely Deployed Wireless Networks," in *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3841-3849, November 2012.

[5] I. Jamil, L. Cariou and J. Hélard, "Novel learning-based spatial reuse optimization in dense WLAN deployments" in *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, Dec., pp. 184, 2016.

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

[7] F. Wilhelmi, S. Barrachina-Muñoz, B. Bellalta, C. Cano, A. Jonsson and G. Neu, "Potential and pitfalls of Multi-Armed Bandits for decentralized Spatial Reuse in WLANs," *Journal of Network and Computer Applications*, vol. 127, 2019, pp. 26-42.

[8] S. Kim, J. Yi, Y. Son, S. Yoo and S. Choi, "Quiet ACK: ACK transmit power control in IEEE 802.11 WLANs," *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Atlanta, GA, 2017, pp. 1-9.

[9] W. Afifi, E. Rantala, E. Tuomaala, S. Choudhury and M. Krunz, "Throughput-fairness tradeoff evaluation for next-generation WLANs with adaptive clear channel assessment," *2016 IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1-6.

[10] I. Roslan, T. Kawasaki, T. Nishiue, Y. Takaki, C. Ohta and H. Tamaki, "Control of transmission power and carrier sense threshold to enhance throughput and fairness for dense WLANs," *2016 International Conference on Information Networking (ICOIN)*, Kota Kinabalu, 2016, pp. 51-56.

[11] S. L. Scott. "A modern bayesian look at the multi-armed bandit," *Applied Stochastic Models in Business and Industry*, 26(6):639-658, 2010.

[12] F. Wilhelmi, B. Bellalta, C. Cano, A. Jonsson, G. Neu, and S. Barrachina-Mu noz. "Collaborative spatial reuse in wireless networks via selfish bandits," *arXiv preprint arXiv:1705.10508*, 2017.

[13] ns-3 (online) available from ⟨https://www.nsnam.org/⟩ (accessed 2019-06-05)

[14] S. Merlin et al., "TGax Simulation Scenarios," doc.: IEEE 802.11-14/0980r16, 2015.