Multi-Armed Bandit Algorithm for Spatial Reuse in WLANs : Minimizing Stations in Starvation

Anthony Bardou[1], Thomas Begin[1], Anthony Busson[1]

Univ. Lyon, ENS de Lyon, UCBL, CNRS, LIP, F-69342, LYON Cedex 07, France
{anthony.bardou,thomas.begin,anthony.busson}@ens-lyon.fr

Mots-clés : *reinforcement learning, starvations, spatial reuse, clear channel assessment.*

1 Introduction

Nowadays, access to WLANs is often regarded as a basic service. However, despite its importance, very few WLANs run at their maximum efficiency. Their current deployments often contain a dense number of access points (APs), which can have a major impact on the WLANs' performance because of the listen-before-talk property of 802.11. The recent amendment to the 802.11 standard (802.11ax or Wi-Fi 6) could be a game-changer as it enables WLANs to dynamically modify the transmission power of APs (a.k.a. `TX_PWR`) as well as their CCA (Clear Channel Assessment) threshold (a.k.a. `OBSS/PD`). In this work, we frame the proper tuning of these two parameters as a Multi-Armed Bandit problem, which allows us to derive an efficient and robust data-driven solution using Thompson sampling, an original sampling of WLAN configurations, and a tailor-made reward function assessing their quality.

2 Methods

2.1 Reward Function

In order to converge quickly to an efficient WLAN configuration in the configuration space $C = ([-82, -62] \times [1, 21])^{N_A}$ (with N_A the number of APs in the WLAN), our reinforcement learning agent must be able to assess the quality of a given configuration through a reward function. We propose a reward function R which is based on the stations' (STAs) effective and attainable throughputs (T_i and T_i^A), T^+ and T^- the sets of STAs not in starvation and in starvation (defined by $T_i < \alpha T_i^A$ for a fixed $\alpha \in [0, 1]$) and the number of STAs in the WLAN N_S. It is defined in Equation 1.

$$R(c) = \frac{|T^-| \prod_{j \in |T^-|} \frac{T_j^-}{\alpha T_j^A} + |T^+| \left(N_S + \prod_{j \in |T^+|} \frac{T_j^+}{T_j^A} \right)}{N_S(N_S + 1)} \tag{1}$$

This reward is bounded between 0 and 1 and mainly increases with the reduction of starvations in the WLAN. We suppose that R is sufficiently smooth over the discrete configuration space C, that is : $\exists \delta \in [0, 1], \forall c_i, c_j \in C, |R(c_i) - R(c_j)| \leq \delta ||c_i - c_j||_1$. This property allows us to split our problem into two independent tasks, (i) look for promising configurations in C to fill a reservoir, done by a first agent, the sampler, and (ii) find the best configuration within the reservoir, done by a second agent, the optimizer.

2.2 Sampler and Optimizer

The sampler exploits the smoothness property of the reward to build a mixture of hyperspheres centered on the best configurations found so far. When triggered, the agent samples a configuration on the surface of these hyperspheres and updates its state, as shown in Figure 1.

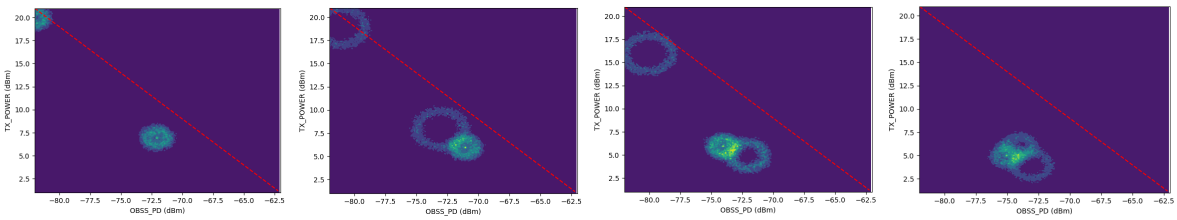
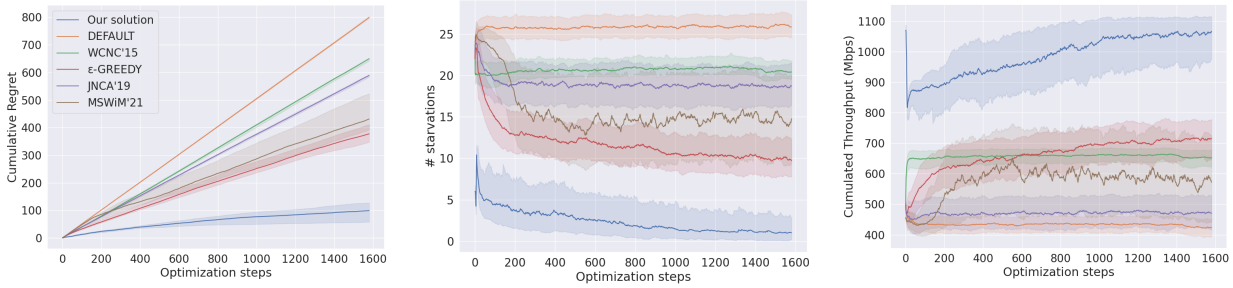


FIG. 1 – Plausible execution of our sampling algorithm for a WLAN with a single AP. The distribution density is shown in colors and the dashed line draws the frontier between authorized and unauthorized configurations according to 802.11ax.

The optimizer is another agent which tries to find $c^* = \text{argmax}_{c \in S} \mathbb{E}[R|c]$, with S the reservoir filled by the sampler when requested. To do so, it tries to approximate the reward distribution (supposed Gaussian) of each configuration c with an iterative Bayesian update of a Normal-Gamma prior with parameters $\left(\hat{\mu}_i, \hat{\lambda}_i, \hat{\alpha}_i, \hat{\beta}_i\right)$. At each step, Thompson sampling uses the beliefs of the agent to select the configuration to test.

3 Results

Our strategy was tested using the realistic network simulator ns-3 [1] and compared with other state-of-the-art solutions such as [2, 3] in full buffer bidirectional (uplink and downlink) traffic on realistic network topologies based on the highly-dense WLAN deployed by Cisco in its offices in San Francisco [4]. Our results show a significant, rapid and lasting improvement of all the performance metrics considered, by factors ranging from 38% to 140% when compared with the default configuration of 802.11 as shown in Figure 2. We can also observe significant improvement when compared to state-of-the-art solutions.



(a) Cumulative Regret (b) Starvations (c) Cumulated Throughput

FIG. 2 – Performance parameters on a realistic topology for each strategy.

Références

[1] *The Network Simulator ns-3*. `https://www.nsnam.org/`. Accessed : 2021-09-30.

[2] F. Wilhelmi et al. "Potential and pitfalls of Multi-Armed Bandits for decentralized Spatial Reuse in WLANs". In : *Journal of Network and Computer Applications* 127 (2019).

[3] M. Shahwaiz Afaqui et al. "Evaluation of dynamic sensitivity control algorithm for IEEE 802.11ax". In : *IEEE Wireless Communications and Networking Conference.* 2015.

[4] *High Density Wi-Fi Deployments*. `https://documentation.meraki.com/Architectures_and_Best_Practices/Cisco_Meraki_Best_Practice_Design/Best_Practice_Design_-_MR_Wireless/High_Density_Wi-Fi_Deployments`. Accessed : 2021-09-30.