Online Cross-Layer Learning in Heterogeneous Cognitive Radio Networks without CSI

Muhammad Anjum Qureshi, Cem Tekin
Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
qureshi@ee.bilkent.edu.tr, cemtekin@ee.bilkent.edu.tr

***Abstract*—We propose a *contextual multi-armed bandit* (CMAB) model for cross-layer learning in heterogeneous *cognitive radio networks* (CRNs). We consider the scenario where *application adaptive modulation* (AAM) is implemented in the *physical* (PHY) layer for heterogeneous applications in the *application* (APP) layer, each having dynamic *packet error rate* (PER) requirement. We consider the *bit error rate* (BER) constraint as the context to mode selector determined by the PHY layer based on the PER requirement, and propose a learning algorithm that learns the modulation with the highest expected reward online over an unknown dynamic wireless channel without *channel state information* (CSI), where the reward is taken as the *Quality of Service* (QoS) provided by the PHY layer to upper layers. We show numerically that the proposed algorithm's expected cumulative loss with respect to an oracle which knows the channel distribution perfectly grows sublinearly in time, and hence, the average loss asymptotically approaches to zero, which in turn yields optimal performance.**

***Keywords*—BER, SNR, regret, PHY layer, AAM, mode selector, feedback, no CSI.**

I. Introduction

A typical wireless system consists of various layers attached in a protocol stack, where each layer performs a specific network function through limited interaction with the other layers [1]. Generally, each layer optimizes its own parameters locally without considering the parameters of the other layers, which results in a suboptimal solution. This motivates joint optimization across layers referred to as *cross-layer optimization*. While cross-layer optimization violates the traditional layered structure, it provides substantial performance improvement [2].

Many prior works on cross-layer optimization assume complete knowledge of the system dynamics, and propose efficient optimization methods using tools such as convex optimization, Lagrange duality, sophisticated scheduling methods for nonconvex problems and combinatorial optimization [3]. For instance, [4] considers spectral efficiency from an optimization perspective with complete knowledge of wireless channel, and proposes cross-layer solution. However, in practice, wireless channel is highly dynamic due to user mobility, multipath and shadowing. Furthermore, obtaining accurate CSI is expensive in terms of system resources. This motivates us to investigate optimal cross-layer learning in the absence of such information.

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under 3501 Program Grant No. 116E229.

In this paper, we consider learning the optimal transmission parameters through repeated interaction with an unknown environment. In our model, the APP layer serves numerous applications with dynamic PER requirements. For each data frame, the PHY layer calculates the target BER based on the target PER, and then the target BER is provided to mode selector as a context. Then, the transmitter chooses an AAM, which is used to transmit the data frame over an unknown wireless channel. After the transmission is complete, the BER is calculated at the receiver end, and communicated to the mode selector. Based on this information and its previous observations, the mode selector calculates a reward that depends on the achieved BER and the target BER, and adapts its AAM selection strategy to maximize its long-term performance. The selected AAM is communicated back to the transmitter via an error-free channel.

We propose a reinforcement learning model and a learning algorithm for the cross-layer learning problem described above. Specifically, we cast this problem as a CMAB [5], which is a generalization of the multi-armed bandit (MAB) [6]. The goal in this problem is to maximize the cumulative reward (or equivalently minimize the regret) by learning the best actions through a process that involves exploration and exploitation. In the MAB, the reward is a random variable that depends on the chosen action. In the CMAB, the reward also depends on the context (side-information) that is revealed before action selection takes place. Thus, the regret in the MAB is defined with respect to the best fixed action, while the regret in the CMAB is defined with respect to the best sequence of actions given the contexts. In prior works, MAB methods are used for opportunistic spectrum access in CRNs to optimize the performance in unknown and dynamically changing environments [7].

CRNs with heterogeneous applications/users usually require different AAM strategies for each user, since each user has a different QoS requirement [8]. For instance, [9] considers heterogeneous CRNs, and proposes dynamic resource allocation schemes for these. Prior works on adaptive modulation selection consider two different types of block-fading channel models based on the coherence time of channel fades [10]: slow block-fading and fast block-fading. In slow block-fading, channel fades remains constant during the transmission of a data frame [11]. This enables channel state estimation at the receiver, which is used for selecting the right transmission mode for the next data frame. In fast block-fading, channel

fades vary even during the transmission of a single data frame, and change from packet to packet. Hence, channel state estimation is not beneficial in choosing the right transmission mode [12]. Several solutions are proposed for the fast block-fading model, such as [13], which uses joint MAP equalization and channel estimation. In this paper, we assume that the unknown channel is a fast block-fading channel, and aim at learning the optimal context dependent AAM without CSI.

The main contributions of this paper are summarized as follows:

- We consider a cross-layer learning problem in a fast block-fading channel, where the current channel condition cannot be accurately observed. Then, we propose a learning algorithm that learns the best QoS dependent AAM by solely using the past BER observations and target BER requirements provided by the PHY layer.
- We compare the performance of the proposed algorithm with an oracle that always chooses the best QoS dependent AAM using perfect knowledge of the channel distribution. As the performance measure, we use the regret, and show via experiments that the regret of the proposed algorithm increases sublinearly over time.

II. System Model and Problem Formulation

A. System Model

The system model is shown in Fig 1. There are three layers in the stack: the PHY layer, the *media access control* (MAC) layer, and the APP layer. The APP layer serves multiple applications sequentially over time. Each application has a dynamic PER constraint, which is used to determine the target BER at the PHY layer denoted by BER^{target}. We also refer to this as the *context*. The conversion from PER to BER is given in [4] for uncoded QAM modulations, which is dependent on the application in hand and error correction algorithm, for instance *forward error coding* (FEC) at the PHY layer or *automatic repeat request* (ARQ) at the MAC layer. When an application runs, it continuously sends its context to the PHY layer. Since there can be multiple applications running at the same time, we order the contexts based on their arrival times, thus in our setting, each context arrival corresponds to a decision epoch. At the PHY layer, the data is transmitted frame by frame through an unknown channel. Each frame may contain multiple packets from the MAC layer. The PHY layer adapts its modulation based on the application and its context. Hence, we call the modulation chosen by the PHY layer for the frame that corresponds to a particular context as the *application adaptive modulation* (AAM).

We consider a very general channel model and assume that neither the channel statistics nor the CSI is available. Thus, the system aims at learning the best AAM on average for each context, where the best AAM is the one that maximizes the expected *bits per symbol* (BPS) rate conditioned on having a BER lower than the BER constraint of the corresponding application.

After the transmission of a data packet is complete, the receiver calculates BER and communicates it to the mode selector. We assume that perfect CRC-based error detection is used at the receiver end via reliable codes, and hence, this computation is error-free. Since the mode selector is at the receiver end, it calculates the instantaneous reward of the packet based on the BER. At the end of each data frame, the mode selector updates the expected reward of the chosen AAM using the average rewards of packets inside the transmitted data frame. Then, the mode selector observes the next context, calculates the estimated best AAM and feeds it back to the transmitter via a fast link feedback channel, after which, the transmitter selects the fed back AAM for the next data frame.

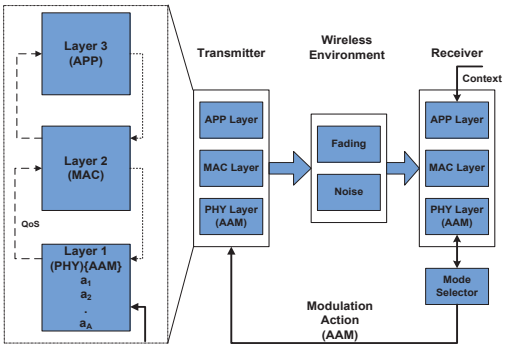

Fig. 1: System Model

B. Action Space

Let t denote the transmission time of the tth data frame and t_p denote the transmission time of the t_pth data packet. At each time t PHY layer chooses an AAM from its AAM set $\mathcal{A} := \{a_1, \ldots, a_A\}$, where A is the number of AAMs. In our setup, AAM a_i corresponds to uncoded QAM modulation with constellation size 4^i, and $A = 5$. The BPS rate of AAM a is denoted by R_a, which is equal to $2i$ for $a = a_i$. Quality of the channel is represented by its *signal-to-noise ratio* (SNR). At the MAC layer, each packet contains N_P bits. At the PHY layer, AAM a maps each packet to a symbol-block containing N_P/R_a symbols. Multiple such blocks constitute one frame containing N_F symbols. The number of symbol blocks in a data frame varies for each AAM $a \in \mathcal{A}$, and is calculated as $N_b^a = N_F R_a / N_P$. This also corresponds to the number of data packets in the data frame. N_P/R_a and $N_F R_a/N_P$ are assumed to be integers.

C. Reward Structure

We consider fixed transmission power and a fast block-fading channel, where the channel fades are considered to be nearly the same as the packet length, and hence, the instantaneous received SNR Γ remains constant during the transmission of a packet. We assume that Γ comes from an unknown distribution p_Γ, and is independently sampled at each packet transmission time t_p. While typically SNR ranges from 0 to 50dB, for technical analysis we linearly rescale the SNR such that it lies in $[0, 1]$.

Let $N(t_p)$ be the number of bits of packet t_p received in error, when the instantaneous SNR is $\gamma(t_p)$ under the selected

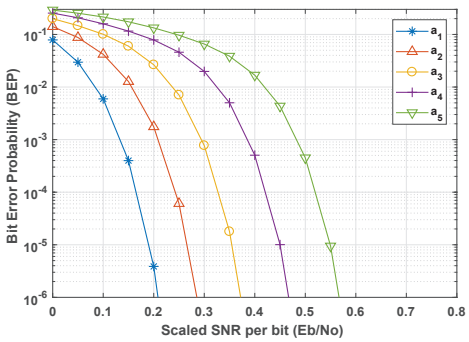

Fig. 2: Bit Error Probability vs SNR for AWGN channel

AAM. The BER for the t_pth transmitted packet is calculated by the receiver as $BER(t_p) = N(t_p)/N_P$. Similarly, let $N_{a,\Gamma}$ denote the number of bits of an N_P bit packet received in error and $BER_{a,\Gamma} = N_{a,\Gamma}/N_P$ denote the BER of AAM a given instantaneous SNR Γ. We have $BER_{a,\Gamma} \in \mathcal{W} := \{0/N_P, 1/N_P \ldots, N_P/N_P\}$. Also, let $\overline{BER}_{a,\gamma}$ denote the *bit error probability* (BEP) of AAM a when $\Gamma = \gamma$. It is known for a wide class of SNR distributions (including Gaussian, Nakagami-m, Rayleigh, Rician) that the BEP monotonically decreases with SNR. Moreover, as shown in Fig. 2, for a fixed SNR the BEP increases when a higher order modulation is selected. For a fixed SNR, the BEP is the expectation of the BER. However, we cannot use BEP curves in Fig. 2 since both the SNR and its distribution are unknown. The (random) reward of AAM a given target BER w is

$$r_a(w) = \begin{cases} R_a/R_{\max} & BER_{a,\Gamma} \leq w \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $R_{\max} = \max_{a\in\mathcal{A}} R_a$ is the maximum BPS rate. This normalization allows us to bound the reward within $[0, 1]$. The expected reward of AAM a for target BER w is given as

$$\mu_a(w) = \mathbb{E}[r_a(w)] = \frac{R_a}{R_{\max}} F_a(w). \quad (2)$$

where $F_a(w) = \Pr(BER_{a,\Gamma} \leq w)$ is the CDF of $BER_{a,\Gamma}$.

It is assumed that the PHY layer provides the target BER w to mode selector from set $\mathcal{W}^{target} := \{w \in \mathcal{W} : w \geq BER^{target}_{\min}\} \subset [0, 1]$, where $BER^{target}_{\min}$ denotes the minimum target BER. We assume that F_a satisfies the similarity information with respect to $w \in \mathcal{W}^{target}$ for all $a \in \mathcal{A}$, which is stated in the following assumption.

Assumption 1. *$\exists L > 0$, such that $\forall a \in \mathcal{A}$, $w_c, w_d \in \mathcal{W}^{target}$ we have*

$$|F_a(w_d) - F_a(w_c)| \leq L|w_d - w_c|.$$

Next, we show that this assumption holds for an example channel model, where $N_P = 1080$, $BER^{target}_{\min} = \frac{2}{1080}$ and the distribution of Γ is given as $p_\Gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp(-\frac{\gamma}{\bar{\gamma}})$, where $\bar{\gamma} := 1/5$ is the average SNR. F_a, $a \in \mathcal{A}$, for this example are given in Fig. 3(i). For this case, it is observed that $L = 21$ satisfies Assumption 1. In addition, $\mu_a(w)$, $a \in \mathcal{A}$ also satisfies Assumption 1.

Algorithm 1 Application Adaptive Modulation (AAM)

1: **Input:** $\mathcal{A}, T, R_a\ \forall a \in \mathcal{A}$
2: **Initialize:** Partition the context set $[0, 1] \supset \mathcal{W}^{target}$ into m_T equal length intervals denoted by $\mathcal{P}_T$
3: $T_{p,a} = 0$, $\hat{\mu}_{p,a} = 0$, $\forall a \in \mathcal{A}, \forall p \in \mathcal{P}_T$
4: $t = 1$, $t_p = 1, h(0) = 0$
5: $R_{\max} = \max_{a\in\mathcal{A}} R_a, A = |\mathcal{A}|$
6: **while** $t \geq 1$
7: Observe $w(t) = BER^{target}(t)$
8: Find a set $p(t)$ in $\mathcal{P}_T$ that contains $w(t)$
9: $\bar{\mu}_{p(t),a} = \hat{\mu}_{p(t),a} + \sqrt{\frac{2(1+2\log(2Am_TT^{3/2}))}{T_{p(t),a}}}$, $\forall a \in \mathcal{A}$
10: $a(t) = \arg\max_{a\in\mathcal{A}} \bar{\mu}_{p(t),a}$
11: $h(t) = h(t-1) + N_b^{a(t)}$
12: $r = 0, \tau = 0$
13: **while** $t_p \leq h(t)$
14: Transmit packet t_p using AAM $a(t)$
15: Observe $BER(t_p)$
16: $r_p = \mathrm{I}(BER(t_p) \leq w(t))R_{a(t)}/R_{\max}$
17: $r = (r\tau + r_p)/(\tau + 1)$
18: $\tau \leftarrow \tau + 1, t_p \leftarrow t_p + 1$
19: **end while**
20: $\hat{\mu}_{p(t),a(t)} = (\hat{\mu}_{p(t),a(t)}\ T_{p(t),a(t)} + r)/(T_{p(t),a(t)} + 1)$
21: $T_{p(t),a(t)} = T_{p(t),a(t)} + 1$
22: $t \leftarrow t + 1$
23: **end while**

D. Regret of Learning

We denote the target BER at time t with $w(t)$. The optimal AAM at time t is $a^*(t) = \arg\max_{a\in\{a_1,\ldots,a_A\}} \mu_a(w(t))$. Computing $a^*(t)$ requires knowledge of p_Γ. In our case, it is impossible to learn p_Γ since there is no CSI. Nevertheless, we compare our algorithm with an oracle that always selects the optimal AAM. We define the performance loss of our algorithm with respect to this oracle as the expected regret, which is given as

$$\mathbb{E}[\mathrm{Reg}(T)] := \sum_{t=1}^{T} \mu_{a^*(t)}(w(t)) - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{a(t)}(w(t))\right]. \quad (3)$$

III. AAM Algorithm

The proposed algorithm (Algorithm 1) is based on a contextual bandit algorithm [14], which uniformly partitions $[0, 1] \supset \mathcal{W}^{target}$ into m_T equal length intervals. This partition is denoted by $\mathcal{P}_T$. The algorithm keeps and updates two parameters for each $a \in \mathcal{A}$ and $p \in \mathcal{P}_T$: (i) $T_{p,a}$ which is the number of times AAM a is selected for contexts in p, and (ii) $\hat{\mu}_{p,a}$ which is the sample mean of the rewards that corresponds to times when AAM a is selected for contexts in p. At each time t, the algorithm identifies $p(t) \in \mathcal{P}_T$ which contains $w(t)$ (if there are multiple such sets, then one of them is randomly selected), and then, chooses the AAM $a(t)$ that maximizes $\bar{\mu}_{p(t),a}$, which is the sum of the sample mean reward $\hat{\mu}_{p(t),a}$ and an uncertainty term given in line 8 of Algorithm 1. This

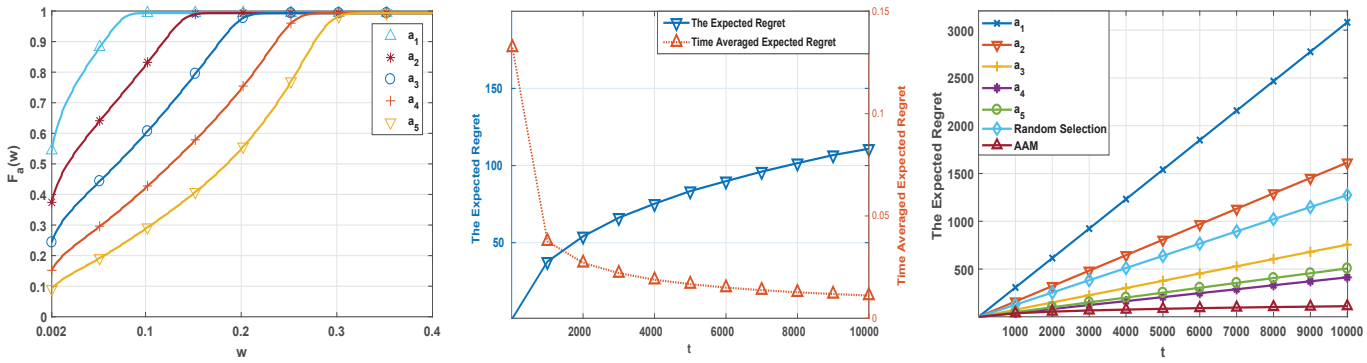

Fig. 3: (i) $F_a(w)$ for the given channel model (ii) The expected regret and the time average expected regret of AAM (iii) Expected regrets of the fixed modulation selection, random selection and AAM

way, $\bar{\mu}_{p,a} + L/m_T$ forms an *upper confidence bound* (UCB) for $\mu_a(w(t))$. This method allows us to exploit the similarity of the AAM rewards given in Assumption 1.

The algorithm also keeps a counter $h(t)$ for the total number of packets to be transmitted up to time t. Since, instantaneous SNR changes from packet to packet, the reward for each packet is first calculated individually using (1), and then, the reward for the tth data frame is obtained by averaging the rewards of the packets inside that frame. Finally, the empirical reward of the chosen AAM is updated.

IV. Illustrative Results

We set $T = 10^4$, $N_P = 1080$, $m_T = \lceil T^{1/3} \rceil$. $w(t)$ takes values in four different intervals that correspond to very low, low, medium and high BER constraints, and is randomly selected from one of these intervals independently from the other times. For simplicity, we assume that the frame that corresponds to AAM a contains exactly $N_b^a = R_a/R_{a_1}$ packets. Hence, for AAM a_1 data frame contains 1 packet, for AAM a_2 data frame contains 2 packets and so on.

The distribution of Γ is given as $p_\Gamma(\gamma) = \frac{1}{\bar{\gamma}} \exp(-\frac{\gamma}{\bar{\gamma}})$, where $\bar{\gamma} := 1/5$ is the average SNR. For packet-level fades, each packet essentially experiences an AWGN channel. The expected regret is calculated based on (3), and reported results correspond to regret averages over 100 runs. In addition, the uncertainty term in the algorithm is scaled with $1/10$ to provide a better exploration and exploitation ratio, which is observed to work well in practice. The total expected regret and the time averaged expected regret are shown in Fig. 3(ii). We also compare the regret of our algorithm with applying a fixed modulation at all times and random selection in Fig. 3(iii). Since AAM exploits contextual information, best action varies for different contexts, which results in a substantially lower regret.

V. Conclusion

In this paper, we propose an online algorithm for cross-layer optimization in heterogeneous CRNs. The proposed algorithm learns the expected best transmission strategy given a dynamic BER constraint in an unknown fast block-fading channel. We compare this algorithm with an oracle that knows the channel distribution and always selects the best transmission strategy for each context. Via numerical results, we show that the regret is sublinear in T for an example setup.

References

[1] Mihaela van der Schaar and Sai Shankar N, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *IEEE Wireless Comm.*, vol. 12, no. 4, pp. 50-58, 2005.

[2] Fangwen Fu and Mihaela van der Schaar, "A new systematic framework for autonomous cross-layer optimization," *IEEE Trans. Vehicular Tech.*, vol. 58, no. 4, pp. 1887-1903, 2009.

[3] Xiaojun Lin, Ness B Shroff and Rayadurgam Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas in Comm.*, vol. 24, no. 8, pp. 1452-1463, 2006.

[4] Qingwen Liu, Shengli Zhou and Georgios B Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Comm.*, vol. 3, no. 5, pp. 1746-1755, 2004.

[5] Aleksandrs Slivkins, "Contextual Bandits with Similarity Information," *Journal of Machine Learning Research*, vol. 15, pp. 2533-2568, 2014.

[6] Tze Leung Lai and Herbert Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4-22, 1985.

[7] Cem Tekin and Mingyan Liu, "Online learning in opportunistic spectrum access: A restless bandit approach," *in Proc. IEEE International Conference on Computer Communications (INFOCOM)*, 2011, pp. 2462-2470.

[8] Babatunde Awoyemi, Bodhaswar Maharaj and Attahiru Alfa, "Optimal resource allocation solutions for heterogeneous cognitive radio networks," *Digital Communications and Networks*, vol. 3, no. 2, pp. 129-139, 2017.

[9] Renchao Xie, F Richard Yu and Hong Ji, "Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing," *IEEE Trans. Vehicular Tech.*, vol. 61, no. 2, pp. 770-780, 2012.

[10] Yu Cao and Steven D Blostein, "Cross-layer optimization of rateless coding over wireless fading channels," *in Proc. 25th IEEE Biennial Symposium on Communications (QBSC)*, 2010, pp. 144-149.

[11] Mohamed-Slim Alouini and Andrea J Goldsmith, "Adaptive modulation over Nakagami fading channels," *Wireless Personal Communications*, vol. 13, no. 1–2, pp. 119-143, 2000.

[12] Pritam Mukherjee and Sennur Ulukus, "Fading wiretap channel with no CSI anywhere," *in Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013, pp. 1347-1351.

[13] Linda M Davis, Iain B Collings and Peter Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *IEEE Trans. on Comm.*, vol. 49, no. 12, pp. 2106-2114, 2001.

[14] Cem Tekin and Mihaela van der Schaar, "Active learning in context-driven stream mining with an application to image mining," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3666-3679, 2015.