OPPORTUNISTIC SPECTRUM ACCESS VIA GOOD ARM IDENTIFICATION

Zhiyang Wang *Ziyu Ying* *Cong Shen*

University of Science and Technology of China

ABSTRACT

In this work, we promote a different tool of multi-armed bandits (MAB), called *arm identification*, to choose a suitable channel for Opportunistic Spectrum Access (OSA) with proven accuracy while satisfying stringent constraints on delay, energy consumption, and channel switches. Noting that finding the *best* channel may not always be the optimal choice, we deviate from the celebrated *best arm identification* framework and adopt *good arm identification (GAI)*, which results in a channel that is "good enough", but requires much less time and energy consumption under the same accuracy requirement. Robustness issues such as delayed or missing feedback are also studied under the new framework. Performance of the proposed algorithm is studied analytically and further corroborated via numerical simulations.

***Index Terms*—** Opportunistic Spectrum Access, multi-armed bandit, arm identification

1. INTRODUCTION

Opportunistic Spectrum Access (OSA) allows secondary users to opportunistically access the channel in a manner that limits interference to the primary users. It can be naturally posed as an *online learning* problem, where the user has to simultaneously exploit the best channels so far and explore the relatively unknown channels. Decision-theoretic approaches are a good fit for this problem [1], including Partially Observable Markov Decision Process (POMDP) [2], and Multi-armed bandits (MAB) [3, 4].

Most existing solutions interlace the sensing and transmitting cycles, with the sensing phase returning an instantaneous estimation of the channel, based on which the user decides which channel to access. The goal therefore is trying to maximize a *long-term* performance metric (e.g., throughput) [2, 3, 4]. A critical shortcoming of these approaches is the slow convergence of learning, especially when the game is hard. In practice, however, many applications have stringent delay requirements, or constraint on the energy consumption of exploration, which is necessary for learning but does not directly benefit the data transmission. Last but not the least, interleaved exploration and exploitation results in frequent channel switches, which is a problem in practice as it may lead to unbalanced user experience or interrupted services.

To address these challenges, in this paper, we focus on a different architecture where channel sensing is aggregated and strictly separated from data transmission which was not considered in earlier studies. The time duration of channel sensing can be set to explicitly satisfy the delay, energy consumption, or switching cost constraints. After the sensing period, the user settles into the selected channel and does not switch. The design goal therefore becomes how to identify a statistically good channel within a given time period with low error probability. In the MAB context, this falls into the field of *best arm identification* (a.k.a. *pure exploration*). Our problem corresponds to the fixed-budget setting – identifying the best arm within a given time budget with the highest accuracy [5]. However, unlike best arm identification, our application does not have to insist on selecting the *best* channel. This is particularly critical when the MAB game is hard, i.e., the difference between the best channel and other sub-optimal channels is very small. In best arm identification, this is often the bottleneck of the sampling complexity, with significant time resource consumption. For OSA, however, there is very little practical difference between these channels, and it is unnecessary to sacrifice significant delay and energy budget on differentiating them.

We thus move away from best arm identification and focus on how to find a "good-enough" channel that satisfies the transmission requirement, such as SINR or QoS. The hope is that by relaxing the objective, we may accelerate the exploration phase and further reduce delay and energy consumption. More specifically, we promote MAB arm identification as a design tool for OSA, and propose a *Good Channel Identification* (GCI) algorithm adapted from a best arm identification algorithm. To further improve the practicality of GCI, we propose a *Any-time Good Channel Identification* (AT-GCI) algorithm that simplifies parameter tuning. Then we address the robustness of AT-GCI with delayed and missing feedbacks by proposing some enhancements. The performances of the proposed strategies are analyzed theoretically with upper bounds on the error probabilities, and further corroborated via numerical simulations using a stochastic geometry model.

2. PROBLEM FORMULATION

We consider an OSA use case where a user is trying to identify a "good-enough" channel among a set $\mathcal{K}$ of K candidate

978-1-7281-1295-4/18/$31.00 ©2018 IEEE

radio channels in the sensing period. The channel quality $r_i(t)$ for channel i at t is measured by the Shannon rate. Based on the measurements, the user estimates the performances of candidate channels and chooses a channel that satisfies the performance requirement. The goal is to design an algorithm that outputs a good channel with high confidence, while satisfying the delay and energy consumption constraints.

This problem can be directly fitted into a stochastic bandit model where each candidate channel corresponds to an *arm*. When channel i is selected at time t, the observation feedback $r_i(t)$ returned to the user as a *reward*. The channel quality is assumed to follow an unknown stationary distribution within the time periods of interest, where $\mu_i = \mathbb{E}[r_i(t)]$ characterizes the mean quality of channel $i \in \mathcal{K}$. The minimum performance threshold for the user is denoted as τ and the goal is to identify a channel whose quality exceeds τ within a given time budget with high accuracy. Specifically, the criterion can be written on the error probability as:

$$e(T) = \mathbb{P}(\mu_{\Omega(T)} < \tau), \tag{1}$$

with $\Omega(T)$ representing the recommended channel index at T. We note that in existing works such as [5, 6], the error probability usually decays *exponentially* with T.

3. CHANNEL IDENTIFICATION ALGORITHMS

3.1. Good Channel Identification

The proposed Good Channel Identification (GCI) algorithm is given in Algorithm 1. At each time step t, the algorithm uses past observations to update the estimated sample mean $\hat{\mu}_i(t), \forall i \in \mathcal{K}$. With the constructed confidence interval $\alpha_i(t)$, the upper and lower bounds on the sample mean are determined: $U_i(t) = \hat{\mu}_i(t) + \alpha_i(t)$, $L_i(t) = \hat{\mu}_i(t) - \alpha_i(t)$. If $\alpha_i(t)$ is constructed properly, $|\hat{\mu}_i(t) - \mu_i| \leq \alpha_i(t)$ with high probability. The algorithm then finds a channel l_t with the minimum upper bound of the gap compared to the empirically optimal channel, which is denoted as $B_i(t) = \max_{k \neq i} U_k(t) - L_i(t)$, and the estimated optimal channel u_t. Then the algorithm selects the most uncertain channel between l_t and u_t, which reveals a new feedback. Finally, at T GCI returns the arm l_t with the smallest index, i.e. $\Omega(T) = \arg\min_{l_t} B_{l_t}(t)$.

The idea of GCI is similar to the UGapEb algorithm in [6], which is designed for the (ϵ, δ) *best* arm identification problem. Since our goal is to select a "good" channel, we note that when ϵ is set as the gap between the optimal mean quality and the threshold, the objective corresponds to criterion (1). With this key observation, the confidence interval is set as $\alpha_i(t) = \sqrt{\frac{T-K}{4\hat{H}_1(t)T_i(t)}}$ with the complexity estimation:

$$\hat{H}_1(t) = \sum_{i=1}^{K} \frac{\mathbb{1}_{\hat{\mu}_i(t) \geq \tau}}{(\hat{\mu}^*(t) - \tau)^2} + \frac{\mathbb{1}_{\hat{\mu}_i(t) < \tau} 4}{(\hat{\mu}^*(t) + \tau - 2\hat{\mu}_i(t))^2}, \tag{2}$$

where $\hat{\mu}^*(t) = \max_{i \in \mathcal{K}} \hat{\mu}_i(t)$.

Algorithm 1 Good Channel Identification

Input: $\mathcal{K}, \tau, T$.
Initialize: Sample each channel once with feedback $\hat{\mu}_i(K)$; Set $T_i(K) = 1$, $t = K + 1$; Calculate $L_i(K)$ and $U_i(K)$.
while $t \leq T$ **do**
  Compute $B_i(t) = \max_{k \neq i} U_k(t) - L_i(t)$.
  $l_t = \arg\min_{i \in \mathcal{K}} B_i(t)$, $u_t = \arg\max_{j \neq l_t} U_j(t)$.
  Select channel $I(t) = \arg\max_{k \in \{u_t, l_t\}} \alpha_k(t)$.
  Observe feedback $r_{I(t)}(t)$ and update $\hat{\mu}_{I(t)}(t) = \frac{\hat{\mu}_{I(t)}(t-1)T_{I(t)}(t-1) + r_{I(t)}(t)}{T_{I(t)}(t-1)+1}$, $T_{I(t)}(t) = T_{I(t)}(t-1)+1$,
  $t = t + 1$.
end while
Return: $\Omega(T) = \arg\min_{l_t} B_{l_t}(t)$.

3.2. Any-time Good Channel Identification

Algorithm 1 requires knowledge of T in advance, which may not always be available in practice. We propose the Any-time Good Channel Identification (AT-GCI) as shown in Algorithm 2, which eliminates this prior knowledge requirement. The main difference lies in that more emphasis is put on the gaps compared to the threshold performance. The estimation indexes are now $|\hat{\mu}_i(t) - \tau| \sqrt{T_i(t)}$, which give more weight to the uncertain channels whose quality is closer to the threshold and those channels with fewer samples. The emphasis on the gaps also leads to improvement compared with the GCI algorithm.

The important feature of this algorithm is that the algorithm can stop at any time while the quality of the output channel is still guaranteed, which makes it suitable for the OSA problem in unknown deployment.

Algorithm 2 Any-time Good Channel Identification

Input: $\mathcal{K}, \tau$ and T.
Initialize: $t = 1, T_i(t) = 0, \hat{\mu}_i(t) = 0 \quad \forall i \in \mathcal{K}$
while $t \leq T$ **do**
  Select channel $I(t) = \arg\min_{i \in \mathcal{K}} |\hat{\mu}_i(t) - \tau| \sqrt{T_i(t)}$.
  Observe reward $r_i(t)$.
  $\hat{\mu}_i(t) = \frac{\hat{\mu}_i(t-1)T_i(t-1) + r_i(t)}{T_i(t-1)+1}$, $T_i(t) = T_i(t-1) + 1$,
  $t = t + 1$.
end while
Return: $\Omega(T) = \max_{i \in \mathcal{K}} \hat{\mu}_i(T)$.

4. PERFORMANCE ANALYSIS

For simplicity, the candidate channels are arranged from 1 to K in a decreasing order of its true quality, and the number of "good enough" channels is m, i.e., $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m \geq \tau > \mu_{m+1} \geq \cdots \geq \mu_K$. $\delta_i = \mu_1 - \mu_i$ and $\Delta_i = |\tau - \mu_i|$

denote the quality gaps compared to the optimal channel and the threshold, respectively.

4.1. Good Channel Identification

Assuming a perfect knowledge of complexity H_1, which can be written as:

$$H_1 = \sum_{i=1}^{K} \max \frac{1}{(\frac{\delta_i + \Delta_1}{2}, \Delta_1)^2} = \frac{m}{\Delta_1^2} + \sum_{i=m+1}^{K} \frac{4}{(\delta_i + \Delta_1)^2},$$

the performance of GCI is similar to that of UGapEb in [6] with ϵ set as $\Delta_1 = \mu_1 - \tau$. More specifically, the following theorem can be proved.

Theorem 1. *With $\hat{H}_1(t)$ set as* (3)*, the error probability $e(T)$ of Algorithm 1 satisfies:*

$$\begin{aligned} e(T) &= \mathbb{P}(\mu_{\Omega(T)} < \tau) = \mathbb{P}(\mu_1 - \mu_{\Omega(T)} > \mu_1 - \tau) \\ &\leq 2KT \exp\left(-\frac{T-K}{2H_1}\right). \end{aligned}$$

4.2. Any-time Good Channel Identification

We first redefine the complexity term as: $H = \sum_{i\in\mathcal{K}} \Delta_i^2 = \sum_{i\in\mathcal{K}} (\mu_i - \tau)^{-2}$, based on which we derive the following main results. The proof is omitted due to space limitations.

Theorem 2. *The error probability $e(T)$ of AT-GCI satisfies:*

$$e(T) = \mathbb{P}(\mu_{\Omega(T)} < \tau) \leq 2T \sum_{i\in\mathcal{K}} e_i(T) \quad (3)$$

$$\leq 2T \sum_{i\in\mathcal{K}} \exp(-2x_i^2), \quad (4)$$

with

$$x_i = \begin{cases} \sqrt{\dfrac{4T}{H}}, & i = \Omega(T) \\ \sqrt{\dfrac{T\left(\frac{4}{H} - \Delta_{i,p}^2\right)}{\Delta_i^2 H}}, & \mu_i \geq \tau, i \neq p, \\ \sqrt{\dfrac{T(1 - 2\Delta_p^2 H)}{2H}}, & \mu_i < \tau, \end{cases}$$

where $\Delta_{j,p} = |\mu_j - \mu_p|$.

5. ROBUSTNESS ISSUES

In this section, we consider the robustness issues in a real-world deployment, including *delayed* or *missing* feedback. We denote the identification algorithms that work without delay, such as GCI and AT-GCI, as $BASE$. In the following, we will show that black-box frameworks can handle these problems with the theoretical performance bounds proposed based on the error probabilities of $BASE$.

5.1. Delayed Feedbacks

With delayed feedback, at time slot t, the user can receive a feedback set $F_t = \{(s, r_i(s)) : s \leq t, s + d_s^i = t\}$, where the immediate quality feedback at time s of channel i is only available after a delay of d_s^i. The Black-box Identification with Delayed Feedback (BI-DF) algorithm provides a solution for this problem. We use $S_i(t)$ to denote the number of feedback the user actually receives from channel i until time t, i.e. $S_i(t) = \sum_{s=1}^{t} \mathbb{1}\{I(s) = i, s + d_s^i \leq t\}$, which plays the same role as $T_i(t)$ in $BASE$.

Algorithm 3 Black-box Identification with Delayed Feedback (BI-DF)

Input: $\mathcal{K}$, τ and T; Algorithm $BASE$.
Initialize: $t = 1, S_i(t) = 0, \hat{\mu}_i(t) = 0 \quad \forall i \in \mathcal{K}$.
while $t \leq T$ **do**
  Calculate the required indexes with $S_i(t)$ and $\hat{\mu}_i(t)$, select channel $I(t)$ according to $BASE$.
  Observe the available feedback set F_t.
  Update estimations in $BASE$ with $(s, r_i(s)) \in F_t$.
  $t = t + 1$.
end while
Return: Channel index according to $BASE$.

The following theorem upper bounds the error probability with delayed feedback. Compared with Theorem 1 and 2, the exponential rate is influenced by the number of feedbacks the user actually receives.

Theorem 3. *The error probability of Algorithm 3 satisfies:*

$$e^{DF}(T) \leq 2T \sum_{i\in\mathcal{K}} \left(e_i^{BASE}(T)\right)^{\frac{G_T}{T}},$$

where $G_t = \sum_{i\in\mathcal{K}} S_i(t)$ is defined as the total number of received feedbacks from K channels.

5.2. Missing feedbacks

Due to channel noise, some feedback packets may not be correctly received by the user at all. We assume that the feedback of channel i may be missing at time t with probability p_i, which can be interpreted as the packet loss rate on the feedback channel. A simple modification to the algorithm can be proposed: skip the update if feedback from sampled channel i is missing for time t. The error probability of this black-box strategy can be upper bounded as the following theorem shows.

Theorem 4. *The error probability of missing feedback strategy satisfies:*

$$e^{MF}(T) \leq 2T \sum_{i\in\mathcal{K}} \left(e_i^{BASE}(T)\right)^{1-p_i}.$$

6. NUMERICAL EXPERIMENTS

6.1. Simulation Setup

In this section we present some numerical results that demonstrate the effectiveness of our learning algorithms in several scenarios. To capture the randomness of the channel quality, we adopt the Poisson Point Process (PPP) to model the number of users in the system as time evolves. In a given region with 1000-meter radius, users on each channel $i \in \mathcal{K}$ are distributed uniformly and geometrically based on a homogenous PPP with density λ_i. We assume the transmit power of each user is identically 5 dBm. The location of the newly coming user is assumed to be fixed at the origin in our simulations. The operations are carried out in a time-slotted system and at each time step $t = 1, 2, \cdots, T$, with the number and locations of users generated according to the aforementioned procedure, the interference to the new user can be calculated based on the path-loss model in [7]. The performance feedback is therefore written as:

$$r_i(t) = \log\left(1 + \frac{P}{N_0 + \sum_{k \in \mathcal{N}_i(t)} P_r^k}\right), \tag{5}$$

where $\mathcal{N}_i(t)$ denotes the set of users allocated on channel i at time t and P_r^k is the interference from user k.

6.2. Simulation Results

We first set $K = 8$ channels with λ_i generated uniformly from 1 to 100. T is set as 100, which is a relatively short period indicating a stringent delay requirement. The threshold is set as $\tau = 6.5$ and $\tau = 3$, respectively. Algorithms are compared to a uniform sampling strategy which is an intuitively simple method often adopted in practice. The error probability for these two settings are shown in figure 1. We can see that both GCI and AT-GCI outperform the uniform strategy and can lower the error probability to less than 0.1 within a relatively short time period. Moreover, the three strategies improve significantly when the threshold is lower. The exponentially decreasing behavior is also obvious from the figure.

Next, we consider the scenario where the feedbacks may be received with a fixed delay $d \in \{5, 10\}$. The impact of missing feedbacks is studied with an i.i.d. Bernoulli model with identical missing probability set as $p_i = 0.5, \forall i \in \mathcal{K}$. The three strategies above are inserted into the black-box frameworks proposed in Sec. 5. We can see that error probabilities of three strategies all increase and larger delay leads to more severe increase. It can be seen that the missing feedback impact the three algorithms in terms of the converge rate and an obvious poor performance in the initial period. The simulation results in figure 2 show the relative robustness of AT-GCI compared to the others. More general scenarios can be realized by setting d_t and p_i as specific values.

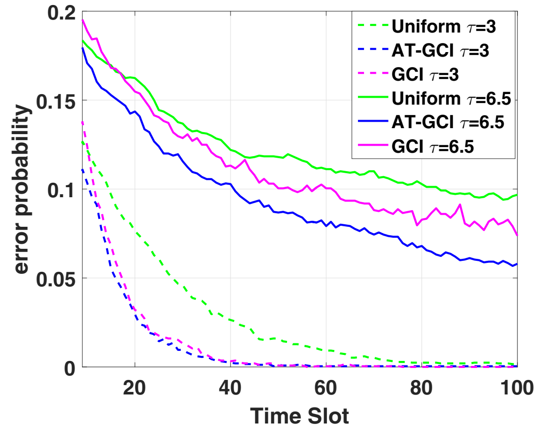

Fig. 1. Error probability when $\tau = 6.5$ and $\tau = 3$.

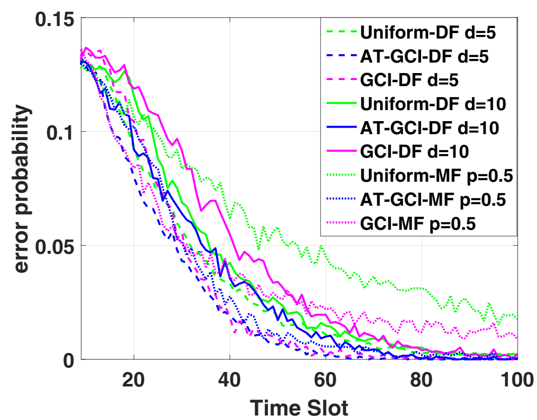

Fig. 2. Error probability when $\tau = 3$, with delay $d = 5$ and $d = 10$, missing probability $p = 0.5$.

7. CONCLUSIONS

Considering OSA applications that call for stringent delay, energy and channel switching constraints, we have studied the channel selection problem with a target of identifying a suitable channel within a finite time horizon with high confidence, instead of trying to maximize the long-term throughput. Furthermore, we have considered the practical limitations of identifying the best performed channel and shifted to the use case where the user only cares about a “good enough” channel. We first propose two algorithms to capitalize the known threshold. We have also improved the algorithms to address some robustness issues in practice. Performance analysis with respect to the selection error probability of these algorithms is carried out. Finally, we verified the performance via numerical simulations.

8. REFERENCES

[1] Q. Zhao and B. M. Sadler, “A survey of dynamic spectrum access,” *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[2] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, April 2007.

[3] W. Jouini, D. Ernst, C. Moy, and J. Palicot, “Upper confidence bound based decision making strategies and dynamic spectrum access,” in *Proc. IEEE Int. Conf. Communications*, May 2010, pp. 1–5.

[4] Cem Tekin and Mingyan Liu, “Approximately optimal adaptive learning in opportunistic spectrum access,” in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1548–1556.

[5] Jean-Yves Audibert and Sébastien Bubeck, “Best Arm Identification in Multi-Armed Bandits,” in *23th Conference on Learning Theory*, June 2010.

[6] Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric, “Best arm identification: A unified approach to fixed budget and fixed confidence,” in *Advances in Neural Information Processing Systems*, 2012, pp. 3212–3220.

[7] 3GPP, “Simulation assumptions and parameters for FDD HeNB RF requirements, R4-092042,” TR 36.814, 2009.