

# Rate and Channel Adaptation in Cognitive Radio Networks Under Time-Varying Constraints

Muhammad Anjum Qureshi<sup>1</sup>, Graduate Student Member, IEEE, and Cem Tekin<sup>1</sup>, Senior Member, IEEE

**Abstract**—We consider dynamic rate and channel adaptation in a cognitive radio network serving heterogeneous applications under dynamically varying channel availability and rate constraint. We formalize it as a Bayesian learning problem, and propose a novel learning algorithm, called Volatile Constrained Thompson Sampling (V-CoTS), which considers each rate-channel pair as a two-dimensional action. The set of available actions varies dynamically over time due to variations in primary user activity and rate requirements of the applications served by the users. Our algorithm learns to adapt its rate and opportunistically exploit spectrum holes when the channel conditions are unknown and channel state information is absent, by using acknowledgment only feedback. It uses the monotonicity of the transmission success probability in the transmission rate to optimally tradeoff exploration and exploitation of the actions. Numerical results demonstrate that V-CoTS achieves significant gains in throughput compared to the state-of-the-art methods.

**Index Terms**—Cognitive radio networks, rate and channel adaptation, opportunistic spectrum access, volatile multi-armed bandits.

## I. INTRODUCTION

### A. Dynamic Rate and Channel Adaptation Problem

AS TRADITIONAL frequency allocation enforced by the Federal Communications Commission (FCC) is quite inflexible, cognitive radios are proposed to opportunistically access free spectrum in order to alleviate frequency scarcity. In cognitive radio networks, a secondary user (SU) usually has access to only those resources (e.g., wireless channels) which are free from their predefined owners i.e., primary users (PUs) at a given time. Thus, the SU opportunistically accesses these resources and makes cognitive decisions aiming to maximize its performance (e.g., throughput) [1]. In heterogeneous networks, there is a multitude of channels each offering time-varying transmission opportunities. In addition to spectrum sharing, *rate adaptation* (RA) is also an important problem in wireless networks, where a transmitter has access to a finite set of transmission rates to choose from at each decision epoch. As the optimal transmission rate and the channel conditions dynamically change over time, traditional RA algorithms (e.g., SampleRate, MiRA, etc.) are outperformed by AI-based learning strategies [1]–[5].

Furthermore, serving heterogeneous applications with diverse quality of service (QoS) requirements may result in

different feasible transmission rates set (e.g., minimum transmission rate constraint or maximum transmission rate limit) for each application. Non-identical traffic models are studied in [6], where a high-end application (e.g., uncompressed video) requires higher constellation (e.g., higher than QPSK), low-end applications are limited to lower constellation, and some applications are subject to medium constellation size (e.g., 16-QAM) to achieve good transmission rate as well as good success probability. Given a modulation scheme, different coding schemes (e.g., 1/2, 3/4, etc.) are available for exploration to satisfy the desired QoS requirement. Therefore, every transmission rate is not a candidate to be explored for the search of the optimal transmission rate [2]. Last but not least, the expected rewards over these resources (e.g., transmission rates) are related and follow a predefined structure. One notable example is the monotonicity of the success probability over transmission rates, which indicates that higher transmission rates result in lower transmission success probabilities [5].

In this letter, we consider the SU's resource allocation problem in a highly dynamic heterogeneous scenario, where not only the channel gains are unknown and may belong to a class of different stochastic distributions, but also the resource set is time-varying (i.e., volatile resource set). Our contributions are listed below:

- We cast rate-channel pair selection problem as an online learning problem, and solve it by designing a learning algorithm called Volatile Constrained Thompson Sampling (V-CoTS), which not only cater to the volatility of resource set but also exploits the structure in the success probability over the available transmission rates.
- We consider learning in an unknown environment, where the SU is initially unaware of the channel characteristics. The available channels need neither to be identically distributed nor to have similar fading characteristics.
- We provide experimental results on cognitive communications over dynamically varying resource sets, and demonstrate that the proposed algorithm achieves considerably higher performance when compared to other state-of-the-art methods. We further provide numerical results to demonstrate significant performance gains in the following simplified scenarios: exploiting the monotone structure in the transmission rates for non-volatile channels (see Fig. 5), and when the action set is one-dimensional (see Fig. 6).

We would like to note that the proposed scheme can also be applied in other technologies such as WiFi and 5G, which may not use primary user/secondary user paradigm. For instance, it can be applied in IEEE 802.11 systems with

Manuscript received July 30, 2020; accepted August 5, 2020. Date of publication August 11, 2020; date of current version December 10, 2020. This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 116E229 and in part by the BAGEP Award of the Science Academy. The associate editor coordinating the review of this letter and approving it for publication was C.-K.-W. Wen. (Corresponding author: Muhammad Anjum Qureshi.)

The authors are with the Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Turkey (e-mail: qureshi@ee.bilkent.edu.tr; cemtekin@ee.bilkent.edu.tr).

Digital Object Identifier 10.1109/LCOMM.2020.3015823

1558-2558 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

TABLE I  
COMPARISON OF V-CoTS WITH RELATED WORKS

Algorithm	KL-UCB-U [1]	RATS [2]	G-ORS [3]	MTS [4]	CoTS [5]	CUL [7]	DRS-TS [8]	V-CoTS (This work)
Rate adaptation	✓	✓	✓	✓	✓	✓	✓	✓
Channel (or width/MIMO mode) selection	✓	✓	✓	×	×	×	×	✓
Structure exploitation	✓	×	✓	×	✓	✓	✓	✓
Bayesian learning	×	✓	×	✓	✓	×	✓	✓
Dynamic rate and channel sets	×	×	×	×	×	×	×	✓

time-invariant or time-varying set of channels (e.g., channel frequencies, channel widths, *multiple input multiple output* (MIMO) modes),<sup>1</sup> or in 5G systems with rapidly varying channels, where a user serving heterogeneous applications targets to maximize its throughput by exploring different transmission rates.

### B. Comparison With Related Works

It is shown in earlier works that resource allocation in rapidly varying wireless networks can be formulated as a multi-armed bandit (MAB) problem, and that this formulation is asymptotically equivalent to maximizing the number of successfully transmitted packets over a given time horizon [1], [3]–[5], [7], [8]. In traditional MAB formulation [9], [10], a player sequentially selects an arm (action) and observes a random realization of the expected reward from an unknown stochastic distribution. Volatile MAB is an important extension of MAB [11], [12], where the set of the arms vary over time. The performance of MAB algorithms is characterized by the regret, which is defined as the difference of the expected reward accumulated by the algorithm from that of the oracle who knows the expected rewards and chooses the optimal transmission parameters in each round. The goal of the learning algorithm is to achieve small regret with respect to the oracle.

In [1], authors propose a non-contextual frequentist algorithm for rate and channel selection, based on the Kullback-Leibler upper confidence bound (KL-UCB), which exploits the unimodal structure of the expected rewards over rate-channel pairs in a steep throughput scenario. Since, the Bayesian approach is proved to be more efficient [8], [13], authors in [4] utilize Thompson sampling for rate selection in wireless networks. The Bayesian approach is extended to exploit the monotone structure of the success probability in the transmission rates in [5]. Authors in [2] propose RATS for rate adaptation in IEEE 802.11ac MIMO systems to efficiently explore combinations of modulation and coding scheme (MCS), channel width, and MIMO mode triple from a predefined reduced set. In [7], authors presented a contextual version of the transmission rate selection problem under time-varying transmit power using the frequentist approach, by exploiting the structure over both the transmission rates and the transmit powers. Later, authors presented a Bayesian counterpart of the proposed frequentist algorithm in [8].

<sup>1</sup>Implementation of V-CoTS does not depend on the specifics of channel models or transmission rates. Thus, it can be used for variety of transmission rates in IEEE 802.11a/b/g/ad with available channels, and in IEEE 802.11n/ac with MIMO modes where each mode can be treated as separate virtual channel and V-CoTS can utilize the monotone structure in transmission rates for each MIMO mode.

Apart from application of volatile and unimodal MAB in resource allocation over unknown and rapidly varying wireless channels, another related line of work investigates application of MAB algorithms for selecting servers within the context of queuing theory and analyzes the novel notion of queuing regret, which compares the learner's queue performance with that of the best server in hindsight [14]. In addition, dynamic channel allocation for deadline constrained traffic is considered in [15] and [16] by using upper confidence bound (UCB) and Thompson sampling based learning algorithms. Moreover, Thompson sampling is used for reconfigurable antenna state selection under unknown channel statistics in [17] and [18].

Departing from earlier works, in this letter we consider rate-channel pair selection in unknown, time-varying channel conditions, and dynamically varying rate-channel sets. The model we consider captures majority of the challenges posed in cognitive radio networks over the rapidly varying channels, in which heterogeneous applications are served by the SU under time-varying PU activity. We propose V-CoTS, which exploits the problem structure to make sequential cognitive decisions for applications with diverse QoS metrics (see Section III). Performance of V-CoTS significantly exceed that of well-known learning methods in such highly dynamic and volatile communication scenarios (see Section IV). Table I lists the key differences of our work from the earlier works.

## II. PROBLEM FORMULATION

### A. System Model

The SU has  $K$  different rates to choose from, represented by the set  $\mathcal{R} = \{r_1, \dots, r_K\}$ . There are  $C$  channels, represented by the set  $\mathcal{C} = \{c_1, \dots, c_C\}$ . The set of rates  $\mathcal{R}$  is ordered such that  $r_1 < \dots < r_K$ . For  $c \in \mathcal{C}$  and  $r \in \mathcal{R}$ ,  $j_c$  and  $i_r$  represent the indices of channel  $c$  and rate  $r$ , i.e.,  $c_{j_c} = c$  and  $r_{i_r} = r$ . The SU serves heterogeneous applications, and each application has its feasible transmission rate set ranging from the minimum rate to be guaranteed to the maximum allowable transmission rate. We also assume that the number of available channels vary with time, due to PU activity. The SU performs channel sensing at the beginning of each round to determine the set of available channels.

### B. MAB Formulation and Reward Structure

In the MAB formulation, the SU makes decisions sequentially over rounds indexed by  $t \in [T]$ ,<sup>2</sup> where  $T$  represents the time horizon. At the beginning of round  $t$ , the SU observes the application to be served in that round

<sup>2</sup>For an integer  $T$ , we let  $[T] := \{1, \dots, T\}$ .

and its feasible rate set<sup>3</sup>  $\mathcal{R}_t$  and performs channel sensing to determine the set of free channels<sup>4</sup>  $\mathcal{C}_t$ . Then, the SU chooses a rate from  $\mathcal{R}_t$  and channel from  $\mathcal{C}_t$ , and transmits at rate  $r(t) \in \mathcal{R}_t$  bits over  $c(t) \in \mathcal{C}_t$ . At the end of the round, it receives ACK/NACK feedback  $x_{r(t),c(t)}(t)$ , which indicates that whether the packet transmission was successful or not, and then, collects the reward as  $r(t) \cdot x_{r(t),c(t)}(t)$ . Here,  $x_{r,c}(t)$  is a Bernoulli random variable with expected value  $\psi_{r,c}$  that takes value 1 if the transmission given rate-channel pair  $(r, c)$  in round  $t$  is successful and value 0 otherwise. We call  $\psi_{r,c}$  the transmission success probability and  $\mu_{r,c} = r \cdot \psi_{r,c}$  the throughput associated with rate-channel pair  $(r, c)$  [8]. We call  $\mu_{r,c}/r_K$  the normalized throughput of rate-channel pair  $(r, c)$ .

Transmission success probabilities exhibit a monotonically decreasing structure over the set of rates [5], given as  $\psi_{r_1,c} > \psi_{r_2,c} \dots > \psi_{r_K,c}$ . The optimal rate-channel pair at round  $t$  is denoted by  $(r^*(t), c^*(t)) = \arg\max_{r \in \mathcal{R}_t, c \in \mathcal{C}_t} \mu_{r,c}$ . Without loss of generality we assume that  $(r^*(t), c^*(t))$  is unique for every  $t$ .

### C. Regret Definition

Let  $\mathcal{A}_t = \mathcal{R}_t \times \mathcal{C}_t$  represent the available action set in round  $t$ . For any  $T$  round available action sequence  $\mathcal{A} = \{\mathcal{A}_1, \dots, \mathcal{A}_T\}$ , the expected regret is defined as

$$R_{\mathcal{A}}(T) = \mathbb{E} \left[ \sum_{t=1}^T \left( \mu_{r^*(t), c^*(t)} - \mu_{r(t), c(t)} \right) \middle| \mathcal{A} \right]. \quad (1)$$

It is obvious that a good learning algorithm should incur small regret. Our goal is to design a learning algorithm for the SU that minimizes the growth rate of the regret, which accounts for maximizing the cumulative throughput.

## III. THE LEARNING ALGORITHM

V-CoTS takes into account the monotone structure of the success probability in transmission rates and time-varying availability of rate-channel pairs while minimizing the regret (pseudocode is given in Algorithm 1). Its main novelty lies in optimizing channel selections together with rate selections under dynamically varying action sets. It is important to note that the SU does not have any control over the available action set, which makes efficient learning much more challenging compared to prior works. In addition, V-CoTS also utilizes monotonicity of  $\psi_{r,c}$  in rates to learn fast [5].

V-CoTS keeps a posterior distribution  $\pi_{r,c} = \text{Beta}(1 + S_{r,c}, 1 + N_{r,c} - S_{r,c})$  of  $\psi_{r,c}$  for each  $(r, c)$ , where  $N_{r,c}(t)$  represents the number of times rate-channel pair  $(r, c)$  was selected before round  $t$ , and  $S_{r,c}(t)$  represents the number of times transmission was successful in rounds where  $(r, c)$  was selected before round  $t$ . These distributions are used to form sample estimates  $\phi_{r,c}$  of  $\psi_{r,c}$  for each  $(r, c) \in \mathcal{A}_t$ . When forming these samples, V-CoTS ensures that the samples are consistent with the monotone structure of  $\psi_{r,c}$  in rates. Let  $\Phi_{t,c} = \{(\phi_{r_{m_t},c}, \dots, \phi_{r_{n_t},c}) \mid \phi_{r,c} \geq \phi_{r',c}, \forall r < r', r, r' \in \mathcal{R}_t\}$  represent the set of samples for channel  $c$  that satisfy

<sup>3</sup>The feasible set at round  $t$  is represented as  $\mathcal{R}_t = \{r_{m_t}, r_{m_t+1}, \dots, r_{n_t}\}$ , where  $m_t \leq n_t$  and  $m_t, n_t \in [K]$ , and  $m_t$  represents the minimum rate to be guaranteed and  $n_t$  represents the maximum allowable transmission rate in round  $t$ . Our algorithm will also work even when  $\mathcal{R}_t$  is an arbitrary subset of  $\mathcal{R}$ .

<sup>4</sup>When channels are non-volatile, then  $\mathcal{C}_t = \mathcal{C}, \forall t \leq T$ .

### Algorithm 1 V-CoTS

---

```

1: Input:  $K, C$ 
2: Initialize:  $t = 1$ 
3: Counters:  $S_{r,c} = 0, N_{r,c} = 0, \forall r \in \mathcal{R}, \forall c \in \mathcal{C}$ 
4: while  $t \geq 1$  do
5:   Observe available channels  $\mathcal{C}_t$ 
6:   Observe feasible rate set  $\mathcal{R}_t$ 
7:   for  $c \in \mathcal{C}_t$ 
8:     Draw  $\phi_{t,c} \sim \mathbf{1}(\phi_{t,c} \in \Phi_{t,c}) \times \prod_{r \in \mathcal{R}_t} \pi_{r,c}$ 
9:      $\theta_{r,c} = r \cdot \phi_{r,c}, \forall r \in \mathcal{R}_t$ 
10:  end for
11:   $[r(t), c(t)] = \arg\max_{r \in \mathcal{R}_t, c \in \mathcal{C}_t} \theta_{r,c}$ 
12:  Transmit using  $[r(t), c(t)]$ , observe feedback
      $x_{r(t),c(t)}(t)$ 
13:   $S_{r(t),c(t)} = S_{r(t),c(t)} + x_{r(t),c(t)}(t)$ 
14:   $N_{r(t),c(t)} = N_{r(t),c(t)} + 1$ 
15:   $t = t + 1$ 
16: end while
    
```

---

monotonicity in round  $t$ . Basically, for each  $c \in \mathcal{C}_t$ , V-CoTS takes samples  $\phi_{t,c} = (\phi_{r_{m_t},c}, \dots, \phi_{r_{n_t},c})$  such that

$$\phi_{t,c} \sim \mathbf{1}(\phi_{t,c} \in \Phi_{t,c}) \times \prod_{r \in \mathcal{R}_t} \pi_{r,c} \quad (2)$$

where  $\mathbf{1}(\cdot)$  is the indicator function.

The above selection ensures that non-zero probability is assigned only to the samples which belong to the set  $\Phi_{t,c}$ . Sampling from (2) can be achieved by first sampling  $\phi_{t,c} \sim \prod_{r \in \mathcal{R}_t} \pi_{r,c}$ , and then, rejecting the samples until the obtained samples belong to  $\Phi_{t,c}$ .<sup>5</sup> A more efficient sampling method that performs rejection by exploiting the inverse transform property, called *sequential inverse transform sampling*, is discussed in [5].

Obtained sample success probabilities are then multiplied by rates  $r \in \mathcal{R}_t$  to obtain throughput samples denoted by  $\theta_{r,c}$  for all available rates in channel  $c$ . The process is repeated for each available channel  $c \in \mathcal{C}_t$ . Finally, at the end of round  $t$ , the rate-channel pair in  $\mathcal{A}_t$  which has the maximum throughput sample is selected for transmission. At the end of the round, the ACK/NACK feedback  $x_{r(t),c(t)}(t)$  and the throughput  $r(t) \cdot x_{r(t),c(t)}(t)$  are observed. Finally, the posterior distribution of rate-channel pair  $(r(t), c(t))$  is updated using the observations gathered in round  $t$ .

## IV. SIMULATION RESULTS

### A. Setup

We have  $K = 10$  rates and  $C = 9$  channels. Transmission success probabilities for each rate and channel are given in Table II, where the optimal rate for a given channel is shown in bold. The expected reward (throughput) is calculated as the product of success probability and corresponding transmission rate. We consider three types of channels. i) Gradual channel: the optimal rate has success probability greater than 0.5, ii) Lossy channel: the optimal rate has success probability

<sup>5</sup>Any other structure can be incorporated to V-CoTS simply by defining appropriate constraint sets  $\Phi_{t,c}$  and performing rejection sampling. When there is no structure, we can simply set  $\Phi_{t,c} = [0, 1]^{|\mathcal{R}_t|}$ .



TABLE II  
SUCCESS PROBABILITIES OVER RATE-CHANNEL PAIRS

$r_k$ (Mbit/s)	MCS	$\psi_{r_k,c_1}$	$\psi_{r_k,c_2}$	$\psi_{r_k,c_3}$	$\psi_{r_k,c_4}$	$\psi_{r_k,c_5}$	$\psi_{r_k,c_6}$	$\psi_{r_k,c_7}$	$\psi_{r_k,c_8}$	$\psi_{r_k,c_9}$
1386.00	(QPSK,1/2)	0.95	0.85	0.99	0.90	0.85	0.98	0.95	0.80	0.95
1732.50	(QPSK,5/8)	0.90	0.75	0.96	0.80	0.80	0.93	0.85	0.70	0.90
2079.00	(QPSK,3/4)	0.85	0.70	0.93	0.75	0.75	<b>0.87</b>	0.80	0.65	0.85
2772.00	(16-QAM,1/2)	0.75	0.55	<b>0.90</b>	0.65	0.65	0.22	0.65	0.52	0.80
3465.00	(16-QAM,5/8)	0.65	<b>0.46</b>	0.18	0.60	<b>0.60</b>	0.16	0.60	<b>0.45</b>	0.75
4158.00	(16-QAM,3/4)	<b>0.60</b>	0.35	0.13	<b>0.55</b>	0.45	0.12	<b>0.55</b>	0.35	0.70
4504.50	(16-QAM,13/16)	0.45	0.30	0.10	0.45	0.35	0.10	0.45	0.27	0.68
5197.50	(64-QAM,5/8)	0.25	0.20	0.07	0.35	0.15	0.07	0.30	0.20	0.66
6237.00	(64-QAM,3/4)	0.15	0.10	0.04	0.20	0.10	0.04	0.20	0.15	0.64
6756.75	(64-QAM,13/16)	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.10	<b>0.62</b>

less than 0.5, and iii) Steep channel: rates have success probabilities either close to 1 or 0 [3], [5]. All channels used in the simulation are generated by adding perturbation to the three main categories. We normalize the throughput by  $r_K$  to have the expected rewards in  $[0, 1]$ . We set  $T = 2.5 \times 10^4$  and present results by averaging over 20 runs.

We assume each application served by the SU remains active for a random amount of time, and the lifetime (in number of rounds) of an application is sampled from the uniform distribution in  $[1, T/25]$ . For each application, feasible rate set is also sampled uniformly from 3 categories, i) QPSK and 16-QAM (i.e.,  $\{r_1, \dots, r_7\}$ ) for low-end applications, ii) 16-QAM and 64-QAM for high-end applications, and iii) 16-QAM only for others. For simplicity, we assume channel 1 is available all the time. For the remaining channels, we set the availability probabilities as  $\{0.8, 0.7, 0.6, 0.7, 0.7, 0.6, 0.7, 0.5\}$ , respectively. Let  $p_i$  be the availability probability of channel  $i$ . We model the PU activity on channel  $i$  as follows. First, we sample  $\chi_i \sim \text{Bernoulli}(p_i)$  and  $\tau_i \in \text{Uniform}([1, T/50])$ . If  $\chi_i = 1$ , then channel  $i$  will be available in the next  $\tau_i$  rounds. Else, it will be unavailable in the next  $\tau_i$  rounds. This model generates bursty PU activity similar to the Gilbert-Elliot channel model [19].

### B. Competitor Algorithms

- i) Oracle: Clairvoyant benchmark that always selects the optimal rate-channel pair in each round (not feasible in practice).
- ii) V-TS: A variant of vanilla Thompson Sampling (TS) [20], which runs on the volatile action set  $\mathcal{A}_t$ .
- iii) V-UCB: A variant of UCB1 [21], which runs on the volatile action set  $\mathcal{A}_t$ .
- iv) CoTS: A constrained Thompson sampling benchmark [5] that exploits the monotonicity of the success probability in transmission rates but ignores the volatility of both the channels and the transmission rates.
- v) CV-CoTS: A variant of the proposed algorithm V-CoTS, which exploits the monotonicity of the success probability in transmission rates as well as the volatility of the channels but ignores the volatility of transmission rates.

A non-volatile algorithm updates the success probabilities for an available channel, however, zero reward (throughput) is obtained for an unavailable rate selection (i.e., when QoS is not satisfied), or when the selected channel is busy due to PU activity.

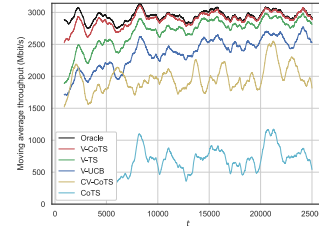


Fig. 1. Moving average throughput: each value at  $t$  is averaged over 20 repetitions of the experiment and previous 900 packet transmissions.

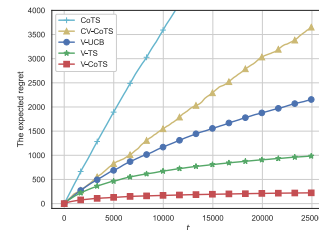


Fig. 2. The expected regret by round  $t$ .

### C. Results

Fig. 1 shows that the optimal (Oracle's) average throughput is around 2944 Mbit/s. V-CoTS follows very closely to this and has average throughput around 2885 Mbit/s. On the other hand, the average throughputs achieved by V-TS, V-UCB, CV-CoTS and CoTS are around 2683, 2362, 1952 and 632 Mbit/s, respectively. The average throughput achieved by V-CoTS is around 98 % of the optimal strategy, and is around 7.5 %, 22 %, 48 % and 356 % higher than that of V-TS, V-UCB, CV-CoTS and CoTS, respectively.

Fig. 2 compares the expected regrets of V-CoTS, V-TS, V-UCB, CV-CoTS and CoTS. It is observed that V-CoTS has the minimum expected regret, and its expected regret at the final round is less than 1/4th of the expected regret of the closest competitor, which is V-TS. This improvement is mainly due to constrained sampling. It is also observed that V-TS outperforms its frequentist counterpart V-UCB by significant margin. It is evident that exploiting only the monotone structure is not sufficient and CoTS performs poorly. A substantial reduction in the regret is achieved by exploiting volatility in the channels in addition to monotonicity in the transmission rates by CV-CoTS, however, this reduction still proves to be non-optimal. Fig. 3 compares the total accumulated number of successfully transmitted Mbits for Oracle, V-CoTS, V-TS, V-UCB, CV-CoTS and CoTS. V-CoTS is the one closest to Oracle. V-TS, V-UCB and CV-CoTS are slow in learning due

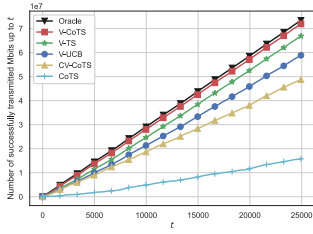


Fig. 3. Number of successfully transmitted Mbits up to round  $t$ .

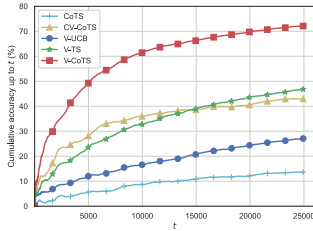


Fig. 4. Accuracy as a function of  $t$ .

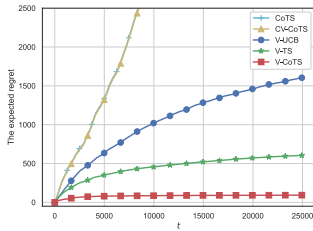


Fig. 5. The expected regret by round  $t$  (non-volatile channels).

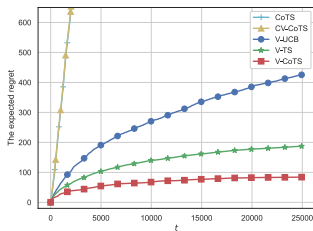


Fig. 6. The expected regret by round  $t$  (one-dimensional actions).

to the fact that they ignore the monotone (or volatile) structure, while CoTS is the worst.

Fig. 4 compares the accuracies of V-CoTS, V-TS, V-UCB, CV-CoTS and CoTS. Accuracy is defined as the fraction of times the selected rate-channel is the optimal one. We observe that by the end, V-CoTS is able to select the optimal rate-channel pair more than 70% of the time, while the closest competitor can only reach around 50%. CV-CoTS and V-TS achieve accuracies close to each other, however, the regret of CV-CoTS is higher than V-TS due to the fact that CV-CoTS receives zero reward when QoS is not satisfied and worst possible regret is incurred in those rounds. Fig. 5 and Fig. 6 provide regret plots for non-volatile channels ( $\mathcal{C}_t = \mathcal{C} = \{c_1, c_2, c_4, c_5, c_9\}, \forall t \in [T]$ ) and transmission rate only actions space ( $\mathcal{C} = \{c_9\}$ ), respectively. The reason behind the similar regret of CV-CoTS and CoTS is the non-volatility of the channels. It is evident that V-CoTS is able to achieve the minimum expected regret in both these simplified scenarios.

## V. CONCLUSION

In this letter, we proposed a Bayesian learning algorithm (V-CoTS) for dynamic rate and channel selection under unknown channel conditions and with time-varying PU user activity and application rate requirements. V-CoTS exploits the monotonicity of transmission success probability in the rates to sample from dynamically changing action sets, and thereby achieving significant gains in throughput compared to other methods.

## REFERENCES

- [1] R. Combes and A. Proutiere, "Dynamic rate and channel selection in cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 910–921, May 2015.
- [2] H. Qi, Z. Lu, X. Wen, and Z. Lu, "Rate adaptation with Thompson sampling in 802.11ac WLAN," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1888–1892, Oct. 2019.
- [3] R. Combes, J. Ok, A. Proutiere, D. Yun, and Y. Yi, "Optimal rate sampling in 802.11 systems: Theory, design, and implementation," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1145–1158, May 2019.
- [4] H. Gupta, A. Eryilmaz, and R. Srikant, "Low-complexity, low-regret link rate selection in rapidly-varying wireless channels," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2018, pp. 540–548.
- [5] H. Gupta, A. Eryilmaz, and R. Srikant, "Link rate selection using constrained Thompson sampling," in *Proc. IEEE INFOCOM-IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 739–747.
- [6] L. Lu, X. Zhang, R. Funada, C. S. Sum, and H. Harada, "Selection of modulation and coding schemes of single carrier PHY for 802.11ad multi-gigabit mmWave WLAN systems," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2011, pp. 348–352.
- [7] M. A. Qureshi and C. Tekin, "Fast learning for dynamic resource allocation in AI-Enabled radio networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 1, pp. 95–110, Mar. 2020.
- [8] M. A. Qureshi and C. Tekin, "Online Bayesian learning for rate selection in millimeter wave cognitive radio networks," in *Proc. IEEE INFOCOM-IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 1449–1458.
- [9] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. Appl. Math.*, vol. 6, no. 1, pp. 4–22, Mar. 1985.
- [10] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, nos. 3–4, pp. 285–294, Dec. 1933.
- [11] Z. Bnaya, R. Puzis, R. Stern, and A. Felner, "Volatile multi-armed bandits for guaranteed targeted social crawling," in *Proc. Workshops 27th Conf. Artif. Intell.*, 2013, pp. 1–3.
- [12] A. Chatterjee, G. Ghalme, S. Jain, R. Vaish, and Y. Narahari, "Analysis of Thompson sampling for stochastic sleeping bandits," in *Proc. Conf. Uncertainty Artif. Intell.*, 2017, pp. 1–10.
- [13] E. Kaufmann, N. Korda, and R. Munos, "Thompson sampling: An asymptotically optimal finite-time analysis," in *Proc. 23rd Int. Conf. Algorithm. Learn. Theory (ALT)*. Springer-Verlag, 2012, pp. 199–213. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-642-34106-9\\_18](https://link.springer.com/chapter/10.1007/978-3-642-34106-9_18)
- [14] S. Krishnasamy, R. Sen, R. Johari, and S. Shakkottai, "Regret of queueing bandits," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 1669–1677.
- [15] S. Cayci and A. Eryilmaz, "Learning for serving deadline-constrained traffic in multi-channel wireless networks," in *Proc. 15th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2017, pp. 1–8.
- [16] S. Cayci and A. Eryilmaz, "Optimal learning for dynamic coding in deadline-constrained multi-channel networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1043–1054, Jun. 2019.
- [17] N. Gulati and K. R. Dandekar, "Learning state selection for reconfigurable antennas: A multi-armed bandit approach," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1027–1038, Mar. 2014.
- [18] T. Zhao, M. Li, and M. Poloczek, "Fast reconfigurable antenna state selection with hierarchical Thompson sampling," in *Proc. ICC-IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
- [19] C. Tekin and M. Liu, "Online learning of rested and restless bandits," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5588–5611, Aug. 2012.
- [20] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. Conf. Learn. Theory*, 2012, pp. 39.1–39.26.
- [21] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.