Link Rate Selection using Constrained Thompson Sampling

Harsh Gupta
ECE and CSL
UIUC
hgupta10@illinois.edu

Atilla Eryilmaz
ECE
The Ohio State University
eryilmaz.2@osu.edu

R. Srikant
ECE and CSL
UIUC
rsrikant@illinois.edu

***Abstract*—We consider the optimal link rate selection problem in time-varying wireless channels with unknown channel statistics. The aim of optimal link rate selection is to transmit at the optimal rate at each time slot in order to maximize the expected throughput of the wireless channel/link or equivalently minimize the expected regret. Lack of information about channel state or channel statistics necessitates the use of online/sequential learning algorithms to determine the optimal rate. We present an algorithm called CoTS - Constrained Thompson sampling algorithm which improves upon the current state-of-the-art, is fast and is also general in the sense that it can handle several different constraints in the problem with the same algorithm. We also prove an asymptotic lower bound on the expected regret and a high probability large-horizon upper bound on the regret, which show that the regret grows logarithmically with time in an order sense. We also provide numerical results which establish that CoTS significantly outperforms the current state-of-the-art algorithms.**

***Index Terms*—Constrained Thompson sampling, Optimal link rate selection, Regret minimization.**

I. Introduction

Optimal link rate selection is an important problem especially in the context of 802.11 systems and other wireless networking systems (see [1], [2], [3] and [4]). At each time slot, the objective of the problem is to choose from a finite set of transmission rates to identify, as quickly as possible, the optimal rate, i.e., the rate maximizing the expected throughput. Along with 802.11 systems, optimal link rate selection problem is also pertinent in cellular wireless systems, especially with the advent of mmWave technology. With the latest advancements in mmWave technology, there is a greater need for fast learning-based rate adaptation algorithms which can perform well in unknown and time-varying channel conditions with limited feedback (see [3]). In such contexts, the standard sampling-based methods or methods relying on perfect channel state information (CSI) are inefficient, costly and unreliable.

We consider a wireless network operating under some MAC protocol and focus on a particular link (transmitter-receiver pair) in this network. Time is indexed so that consecutive time slots are the time slots at which this link is chosen to transmit. Thus, we effectively consider a single link in this paper and we are interested in choosing the optimal transmission rate for this link. We refer to [1] and [3] for earlier work on this problem, and later we will elaborate on our contributions with respect to these prior works. In particular, we consider a time varying wireless channel/link $(h(t))_{t\geq 0}$. At each time slot t, the channel allows transmission at one of the following n rates of transmission: $r_1, r_2, ..., r_n \in \mathcal{R}$. Without loss of generality, we assume $r_1 < r_2 < ... < r_n$. The corresponding probabilities of success for the transmission at these rates are assumed i.i.d. at each time slot and are given by the vector $\theta = (\theta_1, \theta_2, ..., \theta_n)$. Observe that, at a given time slot t, if a transmission at rate r will be successful in the particular channel state $h(t)$, transmission at all rates less than r will also be successful. Therefore, $1 \geq \theta_1 \geq \theta_2 \geq ... \geq \theta_n \geq 0$. Let Θ denote the set of valid rate success probability vectors, i.e., $\Theta = \{\lambda : 1 \geq \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0\}$. The aim of optimal link rate selection is to transmit at the optimal rate r^* (at each time slot) so that the expected throughput is maximized. Let i^* be the index of r^* in the set $\mathcal{R}$, i.e., $r^* = r_{i^*}$. Mathematically, r^* essentially solves the following optimization problem:

$$r^* = r_{i^*} = \arg\max_{r_i} r_i \theta_i. \tag{1}$$

If the vector θ is known, the above optimization problem can be solved easily. But, in most practical applications, the channel statistics are unknown and hence there is no information on the vector θ. This lack of information necessitates the use of online/sequential learning algorithms, which learn the optimal rate over time by transmitting at various rates and gaining information about their probabilities of success (from the history of transmissions and their outcomes). Such online algorithms encounter an *exploration vs. exploitation* trade-off (see [5], a survey on multi-armed bandit problems), i.e., while they have to explore different rates to gain more accurate information, they also have to simultaneously exploit the information gained to transmit at the best possible rate.

A quantity often used to quantify the performance of online algorithms is expected regret. In order to define expected regret, we will introduce some notation first. Since the model we use is similar to the one used in [1] and [3], we will

This research has been supported in part by NSF grants: NeTS 1718203, CMMI 1562276, ECCS 16-09370, CNS-NeTS-1514260, CNS-NeTS-1717045, CMMI-SMOR-1562065, CNS-ICN-WEN-1719371, and CNS-SpecEES-1824337, the DTRA grants: HDTRA1-15-1-0003, and HDTRA1-18-1-0050 and ARO Grant W911NF-16-1-0259.

use similar notation to make our analysis and results more accessible to a reader familiar with those works. Let $r(t)$ denote the rate of transmission chosen at time slot t. Let $i(t)$ denote the index of $r(t)$ in the set of rates $\mathcal{R}$, i.e., $r(t) = r_{i(t)}$. Let $X(t)$ denote the outcome of the transmission at time slot t, i.e., $X(t) = 1$ in the case of a successful transmission and $X(t) = 0$ otherwise. Note that $X(t)$ is a Bernoulli random variable with parameter $\theta_{i(t)}$. Observe that the optimization problem given by Equation (1) can be rewritten as:

$$r^* = r_{i^*} = \arg\max_{r_i} \mathbb{E}[r(t) \times X(t) | r(t) = r_i, \theta].$$

Expected regret for T time slots is defined as the expected loss in throughput incurred by the algorithm due to transmission at sub-optimal rates. Let $R(T)$ denote the regret for T time slots. Mathematically:

$$\mathbb{E}[R(T)] = \mathbb{E}[\sum_{i=1}^{T} \{r_{i^*}\theta_{i^*} - r_{i(t)}\theta_{i(t)}\}].$$

Let $N_i(T)$ denote the number of times transmission was made at rate r_i until time T. Also, let $\Delta_i = r_{i^*}\theta_{i^*} - r_i\theta_i$ denote the loss in expected throughput because of transmitting at rate r_i instead of rate r_{i^*}. A more useful way to rewrite expected regret is the following:

$$\mathbb{E}[R(T)] = \mathbb{E}[\sum_{i \neq i^*} N_i(T)\Delta_i] = \sum_{i \neq i^*} \mathbb{E}[N_i(T)]\Delta_i. \quad (2)$$

Another quantity which is useful in quantifying the performance of online algorithms is simply the number of times transmissions at sub-optimal rates are made, i.e.:

$$R'(T) = \sum_{i \neq i^*} N_i(T) \quad (3)$$

Note that $R'(T)$ is a random variable. We will study both the expected regret and $R'(T)$ in this paper.

In [1], the authors tackle the optimal link rate selection problem by treating each rate as an independent arm in the standard multi-armed bandit problem setup. Although this approach overcomes certain challenges associated with the problem, it does not exploit the structure in the set Θ as treating the rates as independent arms implies lack of ordering in the components of the vector θ. They take a KL-UCB inspired frequentist approach and present an algorithm called KL-R-UCB, which achieves logarithmic regret. With an additional assumption that the expected throughput at different rates is unimodal, they present an asymptotically optimal algorithm called G-ORS (also, see [6] for a Thompson sampling inspired algorithm for the unimodal case).

In [3], the authors treat the problem similarly and present a Thompson sampling inspired algorithm called Modified Thompson Sampling (MTS). This algorithm takes a Bayesian approach and is shown to have the same regret upper bound as KL-R-UCB, since it also does not exploit the structure in the set Θ. They also provide a lower bound for the problem but only for the very specific case of three channel states and rate $r_1 = 0$. It is worth noting that while both KL-R-UCB and MTS have been shown to have the same regret upper bound, simulations in this paper indicate that MTS performs significantly better than KL-R-UCB. In this paper, we do not need any additional assumptions on Θ such as unimodality since such assumptions are hard to justify in practice. However, our algorithm can easily incorporate any additional structure.

Our main contributions are the following:

1) We have designed an algorithm called Constrained Thompson Sampling (CoTS). CoTS exploits the structure in the set Θ efficiently (i.e., the fact that $\theta_1 \geq \theta_2 \geq ... \geq \theta_n$) and is more general in the sense that any additional structure in Θ (such as unimodality) can also be incorporated with minor tweaks in the same algorithm (unlike previous approaches where different constraints were tackled using very different algorithms). We also present SITS, an efficient and fast way to implement CoTS in practice (see Sections III-A and III-B).
2) We provide theoretical guarantees for the regret achieved by CoTS by proving a high probability large-horizon logarithmic upper bound for the notion of regret quantified by $R'(T)$ (see Section IV).
3) We prove an asymptotic lower bound for the expected regret (given by Equation (2)) achieved by any algorithm for the optimal link rate selection problem. We note that this lower bound is established without any unimodality assumption as in [1] or any assumptions on the number of channel states or rates as in [3] (see Section V).
4) We provide numerical results to establish the superiority of CoTS over the current state-of-the-art and to show that it achieves the theoretical lower bound (see Section VI).

II. Existing Algorithms

In this section we discuss some existing work on the optimal link rate selection problem, before moving on to the next section where we present CoTS.

Several link rate selection algorithms (also known as rate adaptation algorithms) relying on sampling-based approaches have been proposed in the literature (for example, see [7], [8] and [9]). At any time slot, these methods rely on the history of outcomes for transmission at different available rates to determine the optimal rate to transmit at. These algorithms primarily use well-engineered heuristics to strike a balance between exploration and exploitation.

Another class of algorithms which can potentially be used are the ones which rely on measurements quantifying the quality of the channel (for example, see [10], [11] and [12]). If the measurements obtained are accurate then these methods can perform really well, but in several practical scenarios that arise in modern time-varying wireless systems, it is costly to obtain reliable measurements. Hence, the viability of such measurement-based algorithms is unclear.

Our aim is to use ideas from the stochastic optimization field (similar to [1] and [3]) to tackle the exploration vs.

Algorithm 1 KL-R-UCB algorithm

for $t = 1, 2, \ldots, n$: transmit at rate r_t.
for $t = n+1, n+2, \ldots$:

1) Compute the set $\mathcal{I} = \arg\max_i q_i(t)$.
2) Transmit at rate $r_{i(t)}$ where $i(t) \in \mathcal{I}$.

end for

exploitation trade-off in a *theoretically principled* and *optimal* manner, rather than tackling it heuristically. To this end, in [1], the authors present a KL-UCB (a variant of the classical UCB algorithm, see [13] and [14]) inspired algorithm called KL-R-UCB (see Algorithm 1). In KL-R-UCB, at each time slot t, the algorithm computes an index $q_i(t), \forall r_i$ as follows:

$$q_i(t) = \max\{q \in [0, r_i] : n_i(t) D(\frac{\hat{\mu}_i(t)}{r_i}, \frac{q}{r_i}) \leq \log(t) + c \log\log(t)\}$$

where $n_i(t)$ denotes the number of times rate r_i has been transmitted in t time slots, $\hat{\mu}_i(t)$ denotes the empirical average of all the outcomes of those transmissions and $D(x, y)$ denotes the KL divergence between two Bernoulli distributions parametrized by x and y. It is shown in [1] that KL-R-UCB achieves logarithmic regret although it does not exploit the structure in Θ. Making an additional assumption of unimodality of expected throughput, i.e., $r_1\theta_1 \leq r_2\theta_2 \leq ... < r_{i^*}\theta_{i^*} > r_{i^*+1}\theta_{i^*+1} \geq ... \geq r_n\theta_n$, authors in [1] present another KL-UCB inspired algorithm called G-ORS (also see [6] for a similar Thompson sampling inspired algorithm) which is very different from KL-R-UCB and is asymptotically optimal.

In [3], the authors take inspiration from Thompson sampling (see [15] and [16]) and present the Modified Thompson Sampling (MTS) algorithm. At any time slot t, MTS maintains independent beta priors for every individual component of θ, then samples a vector $\lambda(t)$ from the product of these priors and transmits at the rate optimal for the sampled vector (see Algorithm 2). Finally, depending on the outcome, it does a Bayesian update to the prior corresponding to the component of θ having the same index as that of the rate at which the transmission was made. Since MTS considers independent beta priors for every component of θ, the set of valid parameters it explores is $[0, 1]^n$ instead of $\Theta = \{x \in [0, 1]^n : x_1 \geq x_2 \geq ... \geq x_n\}$. It is shown in [3] that MTS also achieves logarithmic regret, similar to KL-R-UCB, since it also does not exploit the fact that the components of θ are non-increasing.

From the above discussion, we observe that there are two major disadvantages associated with the current state-of-the-art algorithms for the optimal link rate selection problem:

1) The current state-of-the-art algorithms such as KL-R-UCB and MTS do not exploit the basic structure in the set Θ, i.e., they do not take advantage of the fact that the probability of success is a non-increasing function of the rate of transmission. If an algorithm can exploit this structure in the problem, it can potentially outperform both KL-R-UCB and MTS.
2) Additional constraints or structure in the set Θ (such as unimodality of the expected throughput) are not handled easily by the current state-of-the-art algorithms. In fact, even for unimodality, there is a completely different set of algorithms. If an algorithm can handle additional constraints more generally, it will be useful in a much wider set of applications and environments.

[1]Beta(a, b), known as the beta distribution is a continuous probability distribution with pdf: $f_{a,b}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, x \in [0, 1]$, $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Algorithm 2 Modified Thompson sampling algorithm

for each rate $r_i, i = 1, 2, ..., n$, set $s_i = 0$ and $f_i = 0$.
for $t = 1, 2, \ldots$:

1) For every rate r_i, draw $\lambda_i(t) \sim$ Beta$(s_i + 1, f_i + 1)$. [1]
2) Compute $i(t) = \arg\max_i r_i\lambda_i(t)$. Transmit at rate $r_{i(t)}$.
3) Observe the random transmission outcome $X(t)$.
4) (Prior Update) If $X(t) = 1$, set $s_{i(t)} = s_{i(t)} + 1$. Else if $X(t) = 0$, set $f_{i(t)} = f_{i(t)} + 1$.

end for

In the next section, we will present CoTS which overcomes the above mentioned disadvantages. CoTS uses the basic structure in Θ to its advantage and at the same time is amenable to several additional constraints in Θ that one might want to incorporate.

III. CoTS: Constrained Thompson Sampling

The reason why KL-R-UCB and MTS do not perform optimally is because they do not exploit the basic structure in the set Θ. Moreover, with additional structure in the set such as unimodality, the performance of these algorithms deteriorates further and one has to come up with different algorithms which are optimal. In the context of these observations, we now present CoTS (see Algorithm 3) and the intuition behind it.

A. Intuition

The idea behind CoTS is intuitive and simple. At each time slot t, we maintain independent beta priors for each component of θ, similar to MTS. But instead of simply sampling from the product of these priors (as in MTS), we sample from a distribution which is proportional to the product of these priors when the value being sampled, say λ, belongs to Θ and is 0 otherwise. Mathematically, we sample from a distribution with the following p.d.f:

$$p_t(\lambda) \propto \mathbb{1}\{\lambda \in \Theta\} \prod_{i=1}^{n} \text{Beta}(s_i(t) + 1, f_i(t) + 1)(\lambda_i).$$

where $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n) \in [0, 1]^n$, and $s_i(t)$ and $f_i(t)$ are the number of successful and failed transmissions respectively until the beginning of the time slot t, for the rate r_i. This simple modification allows us to exploit the structure in the set Θ by assigning non-zero probability only to the parameters which belong to the set Θ.

Algorithm 3 Constrained Thompson sampling algorithm

for each rate $r_i, i = 1, 2, ..., n$, set $s_i = 0$ and $f_i = 0$.
for $t = 1, 2, \ldots$:

1) Draw $\lambda(t) \sim \mathbb{1}\{\lambda(t) \in \Theta\} \times \prod_{i=1}^{n} \text{Beta}(s_i + 1, f_i + 1)$.
2) Compute $i(t) = \arg\max_i r_i \lambda_i(t)$. Transmit at rate $r_{i(t)}$.
3) Observe the random transmission outcome $X(t)$.
4) (Prior Update) If $X(t) = 1$, set $s_{i(t)} = s_{i(t)} + 1$. Else if $X(t) = 0$, set $f_{i(t)} = f_{i(t)} + 1$.

end for

In [3], the authors state that the reason one has to treat different components of θ independently is that it is difficult to come up with an easy-to-update prior for Thompson sampling that incorporates the non-increasing property of the components of θ (or any other structure such as unimodality). In CoTS, we still maintain different beta priors for each component of θ to keep the updates simple, but in order to exploit the structure in the set Θ, we restrict the joint distribution to have non-zero weight only for valid parameters in the set Θ. At any time slot t, let the rate selected for transmission be $r_{i(t)}$ and let the outcome of transmission be $X(t)$. Then, the prior $p_{t+1}(\lambda)$ after the Bayesian update will be:

$$p_{t+1}(\lambda) \propto \mathbb{1}\{\lambda \in \Theta\} \prod_{i=1}^{n} \text{Beta}(s_i(t) + 1, f_i(t) + 1)(\lambda_i) \times \lambda_{i(t)}^{X(t)} (1 - \lambda_{i(t)})^{1 - X(t)},$$

Simplifying the above expression, we get:

$$p_{t+1}(\lambda) \propto \mathbb{1}\{\lambda \in \Theta\} \prod_{i \neq i(t)} \text{Beta}(s_i(t) + 1, f_i(t) + 1)(\lambda_i) \times \text{Beta}(s_{i(t)}(t) + 1 + X(t), f_{i(t)}(t) + 1 + (1 - X(t))). \quad (4)$$

Hence, to update the prior distribution, we need to simply update the number of successes or failures corresponding to the beta prior of the component of θ with the same index as that of the rate transmitted (Step 4 in Algorithm 3). Thus, maintaining different beta priors for every component of θ allows CoTS to have easy prior updates (similar to MTS), whereas the restriction imposed on the joint distribution allows it to exploit the structure in Θ. Observe that CoTS is essentially an exact Thompson sampling algorithm, whereas MTS is not. Therefore, a different prior distribution can also be used in place of the prior distribution used by CoTS as long as its Bayesian update is easy and exact.

Also, note that CoTS is general in the sense that the set Θ can have any additional structure on top of the basic property of non-increasing components (such as unimodality) and the algorithm will still work. The indicator function in the joint prior distribution can incorporate any structure in the set Θ, while keeping the prior updates simple as shown in Equation (4). Therefore, CoTS allows us to overcome both the disadvantages associated with the current state-of-the-art algorithms discussed in the previous section.

B. Efficient Implementation

In this subsection, we will discuss some efficient ways of implementing CoTS. Since the prior update step is straightforward, the main focus for improving the efficiency lies on Step 1, i.e., sampling $\lambda(t)$ from the prior distribution.

1) Rejection Sampling: One straightforward way to implement CoTS is to use rejection sampling, i.e., sample $\lambda(t) \sim \prod_{i=1}^{n} \text{Beta}(s_i + 1, f_i + 1)$ and reject the samples until $\lambda(t) \in \Theta$. The main advantage of rejection sampling is that it is easy to implement. Also, rejection sampling is general in the sense that as long as the operation of checking whether a sampled value lies in Θ can be done efficiently, it does not require any other problem-dependent alterations. But the main disadvantage of rejection sampling is that it can be really slow. For example, if the probability of obtaining a valid parameter $\lambda(t) \in \Theta$ when sampling from the distribution $\prod_{i=1}^{n} \text{Beta}(s_i + 1, f_i + 1)$ is x, then the expected number of times in which one samples a valid parameter is $\frac{1}{x}$. Thus, if x is really small, the expected sampling time is really large. Therefore, we need to have a faster sampling method, especially in the cases where the progress of the algorithm will result in x taking small values.

2) Sequential Inverse Transform Sampling (SITS): For the basic structure in Θ, as well as for unimodality, we present a technique to speed up the sampling step for CoTS, called Sequential Inverse Transform Sampling (SITS). The idea behind SITS is to sample different components of $\lambda(t)$ sequentially (instead of all at once and then rejecting), while simultaneously ensuring that the sampled components satisfy the structure in Θ. For example, consider the basic non-increasing components structure in Θ. We observe that the prior distribution at time t can be written as:

$$p_t(\lambda) \propto \prod_{i=1}^{n} \mathbb{1}\{\lambda_{i-1} \geq \lambda_i\} \text{Beta}(s_i + 1, f_i + 1)$$

where $\lambda_0 = 1$. Therefore, to sample fast, we can simply sample $\lambda_1(t)$ from $\text{Beta}(s_1+1, f_1+1)$, then sample $\lambda_2(t)$ from $\text{Beta}(s_2+1, f_2+1)$ while restricting it to be less than $\lambda_1(t)$ and so on. To sample a random variable Z from $\text{Beta}(x, y)$ quickly while restricting it to lie between interval $[a, b]$ (instead of interval $[0, 1]$), we can use inverse transform sampling (see [17]) as follows:

1) Let F denote the cumulative distribution function of $\text{Beta}(x, y)$. Let $\alpha_0 = F(a)$ and $\alpha_1 = F(b)$.
2) Sample a random variable U uniformly from the interval $[\alpha_0, \alpha_1]$, i.e., $U \sim \mathcal{U}(\alpha_0, \alpha_1)$.
3) $Z = F^{-1}(U)$ is the required random variable.

The above technique speeds up the sampling process and unlike rejection sampling, makes the sampling time independent of the probability of sampling a valid parameter from $\prod_{i=1}^{n} \text{Beta}(s_i + 1, f_i + 1)$. For the case of unimodality, a similar procedure can be followed except that now for every component being sampled, we simply need to ensure that it continues to maintain unimodality along with the non-

increasing property. Since the basic ideas are the same, we will skip the details for using SITS with unimodality.

Remark 1. *One of the main reasons that the authors in [3] assumed independence across different rates to design MTS is that, without this assumption, exact Thompson sampling was deemed to be computationally infeasible. Therefore, one of the key contributions of this paper is the design of an efficient exact Thompson sampling algorithm in the implementation of CoTS using SITS. Also, note that the computational and storage complexity of CoTS (exploiting the basic structure of non-increasing components) is linear in the number of rates, same as that of KL-R-UCB and MTS.*

IV. Upper Bound

In this section, we present theoretical guarantees for the performance of CoTS in terms of a high probability large-horizon upper bound on the number of times transmissions at sub-optimal rates are made. We utilize the results obtained in [16] to this end. As in [16], we make some simplifying assumptions to make the analysis tractable. In particular, we assume that the possible values for Θ lie in a discrete set. We state the assumptions more precisely next.

Let π_t denote the prior at the beginning of time slot t. We make the following assumptions:

Assumption 1 (Finitely many transmission rates). $|\mathcal{R}| < \infty$.
Assumption 2 (Finite Θ and non-zero initial probability on θ). $|\Theta| < \infty$, i.e, $\Theta = \{\zeta^{(1)}, \zeta^{(2)}, ..., \zeta^{(L)}\}$. Moreover, $\theta \in \Theta$ and $\pi_1(\theta) > 0$.
Assumption 3 (Strictly decreasing probability of success). For all $\zeta \in \Theta$, $\zeta_1 > \zeta_2 > ... > \zeta_n$.
Assumption 4 (Unique optimal rate) The optimal transmission rate is unique, i.e., $r_{i^*}\theta_{i^*} > r_i\theta_i, \forall i \neq i^*$. Under the above assumptions, we have the following result:

Theorem 1. *Let $N_i(T)$ denote the number of times a transmission at rate r_i is made until time slot t. Under Assumptions 1-4, a high probability large-horizon upper bound holds for CoTS as follows. For any $\delta, \epsilon \in (0,1)$, $\exists\, T^* \geq 0$, such that $\forall\, T \geq T^*$, with probability at least $1-\delta$, we have:*

$$R'(T) = \sum_{i \neq i^*} N_i(T) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right) \sum_{i \neq i^*} \frac{\log T}{D(\theta_i, \frac{r_{i^*}\theta_{i^*}}{r_i})} + C$$

where $C = C(\delta, \epsilon, \mathcal{R}, \Theta, \pi)$ is a problem-dependent constant independent of T and $D(x,y)$ denotes the KL divergence between two Bernoulli distributions parametrized by x and y respectively.

Proof. Our upper bound analysis uses the main result from [16], which gives a high probability large-horizon upper bound for the number of times a sub-optimal action is played by exact Thompson sampling for a complex online problem. As discussed in the previous section, CoTS is an exact Thompson sampling algorithm for the optimal link rate selection problem and hence the main result from [16] can be used to quantify its performance.

We note that the optimal link rate selection problem is a special case of the general complex online problem setup outlined in [16]. The set of actions $\mathcal{A}$ we have is essentially the set of transmission rates, i.e. $\mathcal{A} = \mathcal{R}$. Also, the observation space is $\mathcal{Y} = \{0, r_1, r_2, ..., r_n\}$, i.e., $\mathcal{Y}$ is the sample space for the possible rewards in terms of throughput. The reward function $h : \mathcal{Y} \to \mathbb{R}$ is the identity function, i.e., reward $z = h(y) = y$. Let $l(y; i, \theta)$ denote the probability of observing $y \in \mathcal{Y}$ when a transmission at rate r_i is made, with the underlying rate success vector θ. For all $y \in \mathcal{Y}$, we have $l(y; i, \theta)$ as follows:

$$l(y; i, \theta) = \begin{cases} \theta_i, & \text{if } y = r_i, \\ 1 - \theta_i, & \text{if } y = 0. \end{cases}$$

Hence, the optimal link rate selection problem is a special case of the general complex online problems considered in [16]. Therefore, the above observation, along with Assumptions 1-4 and the fact that CoTS is an exact Thompson sampling algorithm imply that we can use Theorem 1 from [16] to quantify the performance of CoTS.

Using Theorem 1 from [16], $\forall\ T \geq T^*$, for any $\delta, \epsilon \in (0,1)$, with probability at least $1-\delta$,

$$\sum_{i \neq i^*} N_i(T) \leq B(\log T) + C'(\delta, \epsilon, \mathcal{R}, \Theta, \pi), \tag{5}$$

where $C'(\delta, \epsilon, \mathcal{R}, \Theta, \pi)$ is a problem dependent constant and $B(\log T)$ is given as follows:

$$\begin{aligned} & B(\log T) \coloneqq \\ & \max \sum_{i=1}^{n-1} z_i(a_i) \\ & \text{s. t. } z_i \in \mathbb{Z}_+^{n-1} \times \{0\}, a_i \in \mathcal{R} \setminus \{r_{i^*}\}, \\ & \quad z_k \succeq z_i, z_k(a_i) = z_i(a_i), k \geq i, \\ & \quad \forall 1 \leq j, i \leq n-1 : \\ & \quad \min_{\lambda \in S_{a_i}(\theta)} \sum_{k=1}^{n-1} z_i(a_k) D(\theta_k, \lambda_k) \geq \frac{1+\epsilon}{1-\epsilon} \log T \\ & \quad \min_{\lambda \in S_{a_i}(\theta)} \sum_{k=1}^{n-1} (z_i(a_k) - \mathbb{1}_{\{k=j\}}) D(\theta_k, \lambda_k) < \frac{1+\epsilon}{1-\epsilon} \log T \end{aligned} \tag{6}$$

where $S_{a_i}(\theta)$ is the set of $\lambda \in \Theta$ which are indistinguishable from θ when r_{i^*} is transmitted and for which a_i is the optimal rate of transmission, i.e.:

$$S_{a_i}(\theta) \triangleq \{\lambda \in \Theta : D(\theta_{i^*}, \lambda_{i^*}) = 0 \text{ and } \arg\max_{r_k} r_k \lambda_k = a_i\}$$

The interpretation of the optimization problem given by (6) is as follows: $\{a_k\}_{k=1}^{n-1}$ is the sequence in which the sub-optimal rates are eliminated by CoTS, i.e., first the rate a_1 is eliminated, then the rate a_2 is eliminated and so on. z_i is the vector storing the number of times transmissions at sub-optimal rates have been made, until the time slot when rate a_i is eliminated. Once a rate is eliminated, it is not transmitted again.

Let $h(i)$ denote the index of the rate a_i in the set $\mathcal{R}$, i.e., $a_i = r_{h(i)}$. Now, we will show that regardless of the sequence in which the rates are eliminated, for a rate $a_i = r_{h(i)}$, any feasible z_i should satisfy $z_i(a_i) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\frac{\log T}{D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}})} + 1$. Let's assume on the contrary that $z_i(a_i) > \left(\frac{1+\epsilon}{1-\epsilon}\right)\frac{\log T}{D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}})} + 1$. We will show that z_i cannot a feasible point of the optimization problem (6) because it violates the last constraint. For any $\lambda \in S_{a_i}(\theta)$, $\lambda_{i^*} = \theta_{i^*}$ and $\lambda_{h(i)} \geq \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}} > \theta_{h(i)}$. We have, for $j = i$:

$$\begin{aligned}\sum_{k=1}^{n-1}(z_i(a_k) - \mathbb{1}_{\{k=i\}})D(\theta_k, \lambda_k) &\geq (z_i(a_i) - 1)D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}}) \\ &> \left(\frac{1+\epsilon}{1-\epsilon}\right)\log T\end{aligned}$$

The first inequality follows from the non-negativity of $z_i(a_k), \forall k$ and the fact that $D(x, y) \geq 0, \forall x, y$. The second inequality follows from the assumption $z_i(a_i) > \left(\frac{1+\epsilon}{1-\epsilon}\right)\frac{\log T}{D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}})} + 1$. Therefore, $z_i(a_i) > \left(\frac{1+\epsilon}{1-\epsilon}\right)\frac{\log T}{D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}})} + 1$ cannot be a feasible point of (6) as it violates the last constraint. Hence, $z_i(a_i) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\frac{\log T}{D(\theta_{h(i)}, \frac{r_{i^*}\theta_{i^*}}{r_{h(i)}})} + 1$. Therefore,

$$\sum_{i=1}^{n-1} z_i(a_i) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\sum_{i \neq i^*}\frac{\log T}{D(\theta_i, \frac{r_{i^*}\theta_{i^*}}{r_i})} + n - 1$$

Combining the above inequality with (5) and (6), we get the result. □

V. Lower Bound

In [3], the authors prove a lower bound for the optimal link rate selection problem using a Lai and Robbins style of analysis (see [18]), but only in the special case of three channel states and rate $r_1 = 0$. In this section, we obtain a general lower bound for the problem, i.e., a lower bound obtained without any assumptions on the number of channel states or the rates.

In order to obtain the general lower bound, we will transform the optimal link rate selection problem setup into a controlled Markov chain framework (similar to [1]) and use results from [19] (quantifying the performance of efficient adaptive decision rules in a controlled Markov chain setup). The result is the following:

Theorem 2. *Let $P = \{i_1, i_1+1, ..., i^*, ..., n\}$ denote the set of indices such that for any $i \in P$, $r_i \geq r_{i^*}\theta_{i^*}$. Let $P' = P\backslash\{i^*\}$. Then, for the n-rates optimal link rate selection problem, the lower bound on expected regret (asymptotically) is given by:*

$$\lim_{T\to\infty}\frac{\mathbb{E}[R(T)]}{\log T} \geq \sum_i c_i\Delta_i, i \neq i^*,$$

where $\Delta_i = r_{i^*}\theta_{i^*} - r_i\theta_i$. The constants c_i are defined as follows:

$\forall\ i \in \{1, 2, ..., i^* - 1\}$, the constants c_i are the solution to the following linear program:

$$\begin{aligned}&\min \sum_{i=1}^{i^*-1} c_i(r_{i^*}\theta_{i^*} - r_i\theta_i), \\ &\text{s. t. } \sum_{l=1}^{i} c_l\mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\}D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1, \forall i \in P', \\ &c_i \geq 0, \forall i.\end{aligned} \tag{7}$$

$\forall\ i \in \{i^* + 1, i^* + 2, ..., n\}$, the constants c_i are the solution to the following linear program:

$$\begin{aligned}&\min \sum_{i=i^*+1}^{n} c_i(r_{i^*}\theta_{i^*} - r_i\theta_i), \\ &\text{s. t. } \sum_{l=i^*+1}^{i} c_l\mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\}D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1, \forall i, \\ &c_i \geq 0, \forall i.\end{aligned} \tag{8}$$

where $D(x, y)$ denotes the KL divergence between two Bernoulli distributions parametrized by x and y respectively.

Proof. Our lower bound analysis uses results obtained in [19] which quantify the performance of efficient adaptive decision rules in a controlled Markov chain framework. In order to use these results, we need to transform our problem to a controlled Markov chain framework. We use the same transformation as used in [1]. For the ease of readability of users already familiar with the aforementioned references, we will reproduce the transformation from [1] and use similar notation as found in [19] and [1].

Consider a controlled Markov chain $(X_t)_{t\geq 0}$ on a finite state space $\mathbb{S} = \{0, r_1, r_2, ..., r_n\}$ with control laws given by the set $\mathbb{U} = \{1, 2, ..., n\}$. The control laws are independent of the state of the Markov chain and correspond to the index of the rate of transmission selected, i.e., if the control law i is selected, the same control (selecting rate r_i) is applied regardless of the state of the Markov chain. Let the transition probability for going from any state $x \in \mathbb{S}$ to any state $y \in \mathbb{S}$ be denoted by $p(x, y; i, \theta)$, where i is the control law selected and $\theta \in \Theta$ is the unknown underlying vector parametrizing the transition probabilities (θ corresponds to the transmission rate success probability vector in the original optimal link rate selection problem). For all $x, y \in \mathbb{S}$, consider $p(x, y; i, \theta)$ as follows:

$$p(x, y; i, \theta) = p(y; i, \theta) = \begin{cases}\theta_i, & \text{if } y = r_i, \\ 1 - \theta_i, & \text{if } y = 0.\end{cases}$$

Let the immediate reward $r(x, i)$ be equal to $r_i\theta_i$. Note that for any control law i, its immediate reward $r(x, i)$ is equal to its expected reward and is independent of the state x. Finding efficient adaptive sequential decision making rules in the above controlled Markov chain framework is equivalent to solving the optimal link rate selection problem. Hence,

the above construction makes the optimal link rate selection problem amenable to results in [19].

Now, consider a fixed $\theta \in \Theta$. We define the set $B(\theta)$ to be the set of all bad parameters $\lambda \in \Theta$ such that when i^* is the control law chosen, λ is indistinguishable from θ, but i^* is not the optimal control law under λ:

$$B(\theta) = \{\lambda \in \Theta : \lambda_{i^*} = \theta_{i^*} \text{ and } \max_i r_i\lambda_i > r_{i^*}\lambda_{i^*}\}.$$

Consider sets $B_i(\theta), i = 1, 2, ..., n$, defined as follows:

$$B_i(\theta) = \{\lambda \in B(\theta) : r_i\lambda_i > r_{i^*}\lambda_{i^*}\}.$$

Note that $B(\theta) = \bigcup_i B_i(\theta)$. Also, note that if $r_i < r_{i^*}\theta_{i^*}$, $B_i(\theta) = \phi$. Let $P = \{i : r_i \geq r_{i^*}\theta_{i^*}\}$. Since $r_1 < r_2 < ... < r_n$, $P = \{i_1, ..., n\}$, where $i_1 \leq i^*$ is the smallest index satisfying $r_{i_1} \geq r_{i^*}\theta_{i^*}$. Define $P' = P \setminus \{i^*\}$.

Using Theorem 1 in [19], we know that $\bar{c} = (c_1, c_2, ..., c_{i^*-1}, c_{i^*+1}, ..., c_n)$, i.e., the vector of constants (in our theorem statement) for the lower bound solve the following linear program:

$$\begin{aligned} &\min \sum_i c_i(r_{i^*}\theta_{i^*} - r_i\theta_i), \\ \text{subject to } &\inf_{\lambda \in B_i(\theta)} \sum_{l \neq i^*} c_l D(\theta_l, \lambda_l) \geq 1, \forall i \in P', \qquad (9) \\ &c_i \geq 0, \forall i. \end{aligned}$$

where $D(\theta_l, \lambda_l)$ denotes the KL-divergence between Bernoulli distributions parametrized by the θ_l and λ_l. Now, all that remains to prove is that the above linear program is equivalent to the two linear programs in the theorem statement.

In order to decouple and simplify the above LP, we will focus on simplifying the first constraint. Without loss of generality, consider $i > i^*$. Note that $i \in P'$. Now, we observe the following:

1) Since $\lambda \in B_i(\theta)$, we know that $\lambda_{i^*}r_{i^*} = \theta_{i^*}r_{i^*}$ and also $\lambda_i > \left\{\frac{\lambda_{i^*}r_{i^*}}{r_i} = \frac{\theta_{i^*}r_{i^*}}{r_i}\right\} > \theta_i$. Since $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$, therefore, for any $\lambda \in B_i(\theta)$:

$$\sum_{l \neq i^*} c_l D(\theta_l, \lambda_l) \geq \sum_{l=i^*+1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}).$$

2) Consider $\lambda(\epsilon) \in B_i(\theta)$ such that $\lambda_l(\epsilon) = \theta_l, \forall l \in \{1, 2, ..., i^*\} \cup \{i+1, i+2, ..., n\}$, $\lambda_i(\epsilon) = \frac{r_{i^*}\theta_{i^*}}{r_i} + \epsilon$ and $\lambda_l(\epsilon) = \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\}\{\frac{r_{i^*}\theta_{i^*}}{r_i} + \epsilon\} + \mathbb{1}\{\theta_l > \frac{r_{i^*}\theta_{i^*}}{r_i}\}\theta_l, \forall l \in \{i^*+1, i^*+2, ..., i-1\}$. It can be easily verified that $\lambda(\epsilon) \in B_i(\theta)$. Now, using $\lambda(\epsilon)$, we get:

$$\begin{aligned} &\sum_{l \neq i^*} c_l D(\theta_l, \lambda_l(\epsilon)) \\ &\qquad = \sum_{l=i^*+1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i} + \epsilon). \end{aligned}$$

Therefore:

$$\begin{aligned} &\lim_{\epsilon \to 0} \sum_{l \neq i^*} c_l D(\theta_l, \lambda_l(\epsilon)) \\ &\qquad = \sum_{l=i^*+1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}). \end{aligned}$$

From the above facts, we can conclude that for $i > i^*$, the first constraint in the LP given by Equation (9) is equivalent to:

$$\sum_{l=i^*+1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1. \qquad (10)$$

Similarly, for $i \in P'$ such that $i < i^*$, we can show that the first constraint in the LP given by Equation (9) is equivalent to:

$$\sum_{l=1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1. \qquad (11)$$

Using Equations (10), (11) in the LP given by Equation (9), we get the following simplified LP:

$$\begin{aligned} &\min \sum_i c_i(r_{i^*}\theta_{i^*} - r_i\theta_i), \\ \text{s. t. } &\sum_{l=1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1, \forall i \in P', i < i^*, \\ &\sum_{l=i^*+1}^{i} c_l \mathbb{1}\{\theta_l \leq \frac{r_{i^*}\theta_{i^*}}{r_i}\} D(\theta_l, \frac{r_{i^*}\theta_{i^*}}{r_i}) \geq 1, \forall i > i^*, \\ &c_i \geq 0, \forall i. \end{aligned}$$

The above LP can be very straightforwardly decoupled into two LPs as in the theorem statement, giving us the final result. □

VI. Simulation Results

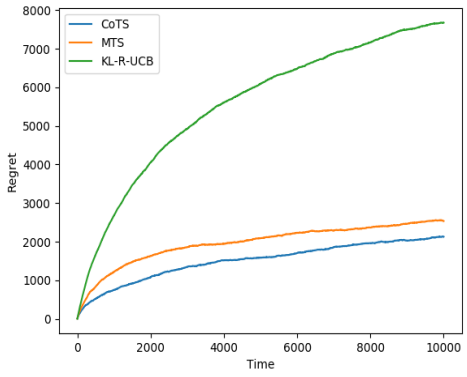
In this section, we present simulation results comparing the performance of CoTS with the current state-of-the-art algorithms. For the optimal link rate selection problem with the basic non-increasing components structure, KL-R-UCB (see [1]) and MTS (see [3]) are the current state-of-the-art algorithms. With the additional constraint of unimodality known, G-ORS has been shown (see [1]) to be asymptotically optimal.

We consider the same experimental setup as in [1], i.e., a single-link 802.11g system with eight available rates from 6 to 54 Mbit/s (also see [7]). The eight rates are as follows (in Mbit/s): $r_1 = 6, r_2 = 9, r_3 = 12, r_4 = 18, r_5 = 24, r_6 = 36, r_7 = 48$ and $r_8 = 54$. We implement all the algorithms in three different scenarios (different values of θ) as used in [7]: *gradual*, *steep* and *lossy*. For all these scenarios, we implement and compare KL-R-UCB, MTS and CoTS (without exploiting unimodality) for the case when only the basic structure in Θ is known. We also implement and compare G-ORS and CoTS (exploiting unimodality), for the case when additional structure of unimodality is known.

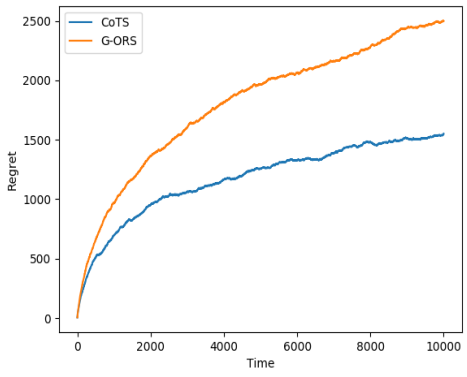
A. Gradual

In the gradual case, we consider rate success probability vector $\theta = (0.95, 0.90, 0.80, 0.65, 0.45, 0.25, 0.15, 0.10)$. Therefore, the vector of expected throughput ξ, i.e., $\xi_i = r_i\theta_i$, is: $\xi = (5.7, 8.1, 9.6, 11.7, 10.8, 9., 7.2, 5.4)$. The defining property of the gradual case is that the optimal rate is the

highest rate with the probability of success greater than 0.5. Figure 1a compares the performance of KL-R-UCB, MTS and CoTS (without exploiting unimodality) for the gradual case. CoTS outperforms both KL-R-UCB and MTS. Another point worth noting is that CoTS and MTS outperform KL-R-UCB by a significant margin. Figure 1b compares the performance of G-ORS and CoTS (exploiting unimodality) for the gradual case. Here again, CoTS outperforms G-ORS. The lower bound constant for the gradual case obtained from Theorem 2 is 526.19, whereas until $t = 10000$, CoTS achieves a constant of 154.78. Although this might seem illogical, we note that the lower bound is asymptotic. It is interesting to note that while it may take a long time to achieve the lower bound, the performance is even better than the lower bound in finite time. This is a feature we have observed in many simulations.



(a) Plot comparing the performance of KL-R-UCB, MTS and CoTS (*without exploiting* unimodality) on the *gradual* case.

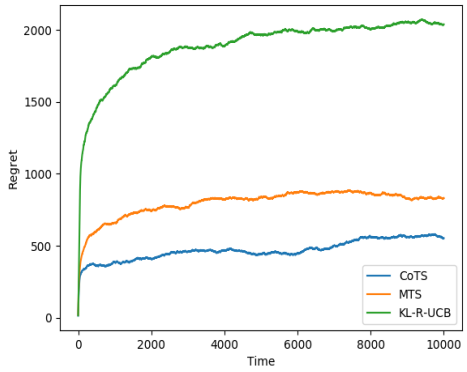


(b) Plot comparing the performance of G-ORS and CoTS (*exploiting* unimodality) on the *gradual* case.

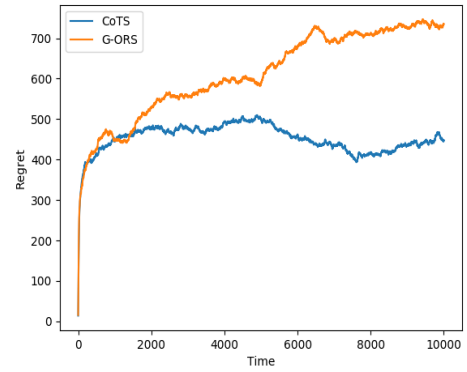
Fig. 1. Performance of CoTS vs. state-of-the-art in 802.11g systems with rate success probabilities characterized by the *gradual* case. Note that CoTS outperforms the current state-of-the-art in both the cases, i.e., whether one exploits unimodality or not.

B. Steep

In the steep case, we consider rate success probability vector $\theta = (0.99, 0.98, 0.96, 0.93, 0.90, 0.10, 0.06, 0.04)$. Therefore, the vector of expected throughput ξ is: $\xi = (5.94, 8.82, 11.52, 16.74, 21.6, 3.6, 2.88, 2.16)$. The defining property of the steep case is that the success probability of every rate is either really high or really low (either close to 1 or close to 0). Figure 2a compares the performance of KL-R-UCB, MTS and CoTS (without exploiting unimodality) for the steep case. Similar to the gradual case, CoTS again outperforms both KL-R-UCB and MTS. Also, CoTS and MTS again outperform KL-R-UCB by a significant margin. Figure 2b compares the performance of G-ORS and CoTS (exploiting unimodality) for the steep case. Here again, CoTS outperforms G-ORS. The lower bound constant for the steep case obtained from Theorem 2 is 45.56, whereas until $t = 10000$, CoTS achieves a constant of 46.49. This shows that CoTS is almost optimal.



(a) Plot comparing the performance of KL-R-UCB, MTS and CoTS (*without exploiting* unimodality) on the *steep* case.



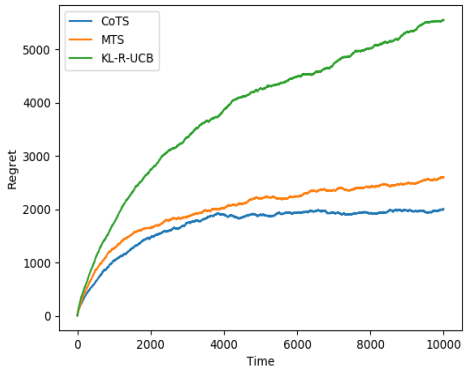
(b) Plot comparing the performance of G-ORS and CoTS (*exploiting* unimodality) on the *steep* case.

Fig. 2. Performance of CoTS vs. state-of-the-art in 802.11g systems with rate success probabilities characterized by the *steep* case. Note that CoTS outperforms the current state-of-the-art in both the cases, i.e., whether one exploits unimodality or not.

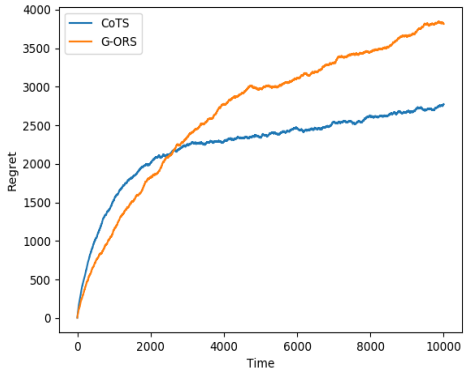
C. Lossy

In the lossy case, we consider rate success probability vector $\theta = (0.90, 0.80, 0.70, 0.55, 0.45, 0.35, 0.20, 0.10)$. Therefore, the vector of expected throughput ξ is: $\xi =$

$(5.4, 7.2, 8.4, 9.9, 10.8, 12.6, 9.6, 5.4)$. The defining property of the lossy case is that the optimal rate has a low success probability, typically less than 0.5, i.e., the system loses significant packets even at the optimal rate. Figure 3a compares the performance of KL-R-UCB, MTS and CoTS (without exploiting unimodality) for the lossy case. Similar to the gradual and steep cases, CoTS again outperforms both KL-R-UCB and MTS. Also, CoTS and MTS again outperform KL-R-UCB by a significant margin. Figure 3b compares the performance of G-ORS and CoTS (exploiting unimodality) for the lossy case. Here again, CoTS outperforms G-ORS. The lower bound constant for the lossy case obtained from Theorem 2 is 401.41, whereas until $t = 10000$, CoTS achieves a constant of 181.44. Again, the performance of CoTS is better than the asymptotic lower bound in finite time.



(a) Plot comparing the performance of KL-R-UCB, MTS and CoTS (*without exploiting* unimodality) on the *lossy* case.



(b) Plot comparing the performance of G-ORS and CoTS (*exploiting* unimodality) on the *lossy* case.

Fig. 3. Performance of CoTS vs. state-of-the-art in 802.11g systems with rate success probabilities characterized by the *lossy* case. Note that CoTS outperforms the current state-of-the-art in both the cases, i.e., whether one exploits unimodality or not.

VII. Conclusion

In this paper, we consider the optimal link rate selection problem in time-varying wireless channels with unknown channel statistics and limited channel state information. We design an algorithm called CoTS which exploits the structure in the problem efficiently and improves upon the current state-of-the-art. We present theoretical upper bounds on its performance and also prove a general lower bound for the optimal link rate selection problem. To corroborate the theory, we present numerical results comparing CoTS with the current state-of-the-art and observe that it performs better across variety of scenarios.

References

[1] R. Combes, A. Proutiere, D. Yun, J. Ok, and Y. Yi, "Optimal rate sampling in 802.11 systems," in *INFOCOM, 2014 Proceedings IEEE (Longer version available on arXiv)*. IEEE, 2014, pp. 2760–2767.

[2] R. Combes, J. Ok, A. Proutiere, D. Yun, and Y. Yi, "Optimal rate sampling in 802.11 systems: Theory, design, and implementation," *IEEE Transactions on Mobile Computing*, 2018.

[3] H. Gupta, A. Eryilmaz, and R. Srikant, "Low-complexity, low-regret link rate selection in rapidly time-varying wireless channels," in *INFOCOM, 2018 Proceedings IEEE*. IEEE, 2018.

[4] R. Combes and A. Proutiere, "Dynamic rate and channel selection in cognitive radio systems." *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 5, pp. 910–921, 2015.

[5] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, pp. 1–122, 2012.

[6] S. Paladino, F. Trovo, M. Restelli, and N. Gatti, "Unimodal thompson sampling for graph-structured arms." in *AAAI*, 2017, pp. 2457–2463.

[7] J. C. Bicket, "Bit-rate selection in wireless networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2005.

[8] A. Kamerman and L. Monteban, "Wavelan®-ii: a high-performance wireless lan for the unlicensed band," *Bell Labs technical journal*, vol. 2, no. 3, pp. 118–133, 1997.

[9] M. Lacage, M. H. Manshaei, and T. Turletti, "Ieee 802.11 rate adaptation: a practical approach," in *Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2004, pp. 126–134.

[10] G. Judd, X. Wang, and P. Steenkiste, "Efficient channel-aware rate adaptation in dynamic environments," in *Proceedings of the 6th international conference on Mobile systems, applications, and services*. ACM, 2008, pp. 118–131.

[11] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive mac protocol for multi-hop wireless networks," in *Proceedings of the 7th annual international conference on Mobile computing and networking*. ACM, 2001, pp. 236–251.

[12] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proceedings of the 8th annual international conference on Mobile computing and networking*. ACM, 2002, pp. 24–35.

[13] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[14] A. Garivier and O. Cappé, "The kl-ucb algorithm for bounded stochastic bandits and beyond," in *Proceedings of the 24th annual Conference On Learning Theory*, 2011, pp. 359–376.

[15] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proceedings of the 25th Annual Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 23. Edinburgh, Scotland: PMLR, 25–27 Jun 2012.

[16] A. Gopalan, S. Mannor, and Y. Mansour, "Thompson sampling for complex online problems," in *Proceedings of the 31st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research. Bejing, China: PMLR, 22–24 Jun 2014, pp. 100–108.

[17] W. contributors, "Inverse transform sampling — Wikipedia, the free encyclopedia," 2018, [Online; accessed 22-July-2018]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Inverse_transform_sampling&oldid=850778210

[18] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[19] T. L. Graves and T. L. Lai, "Asymptotically efficient adaptive choice of control laws incontrolled markov chains," *SIAM journal on control and optimization*, vol. 35, no. 3, pp. 715–743, 1997.