Low-Complexity, Low-Regret Link Rate Selection in Rapidly-Varying Wireless Channels

Harsh Gupta  
ECE and CSL  
UIUC  
hgupta10@illinois.edu

Atilla Eryilmaz  
ECE  
Ohio State University  
eryilmaz@ece.osu.edu

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

Abstract—We consider the problem of transmitting at the optimal rate over a rapidly-varying wireless channel with unknown statistics when channel quality is very limited. One motivation for this problem is that, in emerging wireless networks, the use of mmWave bands means that the channel quality can fluctuate rapidly and thus, one cannot rely on full channel-state feedback to make transmission rate decisions. Inspired by related problems in the context of multi-armed bandits, we consider a well-known algorithm called Thompson sampling to address this problem. However, unlike the traditional multi-armed bandit problem, a direct application of Thompson sampling results in a computational and storage complexity that grows exponentially with time. Therefore, we propose an algorithm called Modified Thompson sampling (MTS), whose computational and storage complexity is simply linear in the number of channel states and which achieves at most logarithmic regret as a function of time when compared to an optimal algorithm which knows the probability distribution of the channel states.

Index Terms—Link Rate Selection, Thompson Sampling, Regret Minimization, Computational Complexity.

I. INTRODUCTION

We are on the verge of an exciting and an unprecedented expansion of the available communication spectrum. In particular, FCC has recently opened up (see [1]) vast spectrum bands (at least 14 GHz in the 57 – 71 GHz range, and more expected) above 28 GHz to public use. These new so-called millimeter Wave (mmW) bands come with their unique dynamics and challenges that demand a fresh look towards the learning and utilization of this new spectrum. On the one hand, the statistical characteristics and sensitivities of these extremely high frequency levels do not fit (see [2]–[5] and references therein for extended discussion) into the commonly used communication radio frequencies (of up to 3 GHz), for which existing cellular technologies and most commonly used 802.11a/b/g/n WiFi protocols are designed. These new channels are highly sensitive to mobility and are subject to drastic time variations that must be accommodated in the learning process.

On the other hand, the vast expansion of the spectrum from previous levels of about 3 GHz by an order of magnitude makes the use of existing estimation and allocation strategies impractical due to the scaling and coordination costs. This motivates us in this work to take a fresh approach to fast learning and resource allocation for multi-rate wireless communication under time-varying and unknown channel conditions.

Traditional communication protocols employ a variety of probing and channel estimation techniques to guide power and rate allocation decisions (see [6]–[8]). While the sophistication and efficiency of these methods vary from carefully engineered cellular technologies to random-access-based WiFi technologies, the common foundation that they are built upon is the assumption that the cost of channel estimation is worth the utility of the acquired channel state information (CSI). This assumption holds in existing systems for two reasons: first, because the channels in the existing communication frequencies are less sensitive to mobilities and thus the CSI can be utilized for a longer duration, and second because the available spectrum of no more than 3 GHz is small enough to track and thus important enough to utilize.

These approaches, however, are not necessarily applicable in the emerging ultra-wideband wireless communication paradigm due to the highly intermittent dynamics and the non-traditional statistics of mmW channels (see [2], [9]–[11]), and the vast scale of the new spectrum (see [1]). In such a setting, where the channel statistics are unknown a priori and the channel conditions are highly time-varying, it is necessary to develop new online learning and adaptive allocation strategies based on limited feedback, such as success/fail signals, that can rapidly converge to optimal solutions with minimal regret.

Several interesting works have explored the learning and rate allocation problem for sum throughput maximization (e.g. [12]) under error bounds (e.g. [13]) based on degraded or ACK/NACK type ARQ feedback. These works, however, do not provide guarantees on short-term performance, such as regret optimality (see [14]–[16]), that are critical in rapidly time-varying channels such as mmW channels.

In this paper, we consider the problem of rate selection for a single user where there is no explicit channel state feedback, but the only feedback available is whether the transmission was successful or not. This problem is related to, but also quite different from, multi-armed bandit problems which have been studied extensively in the context of spectrum sharing in wireless networks (see [17]–[19]).
are in the context of multiple users, somewhat surprisingly, the rate selection problem with limited feedback is challenging even for a single user which is what we focus on in this paper.

Our main contributions in this paper are the following:

- We pose the optimal link rate selection problem so that the general Thompson Sampling (TS) algorithm (see [15]) can be used. However, we identify computational complexity and storage issues with the general TS algorithm which renders it infeasible (see Sections III-A and III-B).
- We design a Modified Thompson Sampling (MTS) algorithm which ignores the fact that a higher transmission rate is less likely to succeed and decouples the rate admissibility probabilities for various transmission rates. Despite this approximation, we show that MTS has logarithmic (or smaller) regret (see Sections III-C and IV).
- We also discuss another way to decouple the rate admissibility probabilities using existing Thompson sampling ideas. However, we show that this approach leads to inferior results compared to our proposed MTS algorithm (see Section IV-A for the theory and Section VI for simulations).
- For a special case, we show that the constant achieved in the logarithmic upper bound for MTS is the tightest possible by obtaining a lower bound using a Lai and Robbins (see [16]) style of analysis (see Section V).
- We conclude the paper with simulation results corroborating the validity of our theoretical guarantees (see Section VI).

II. Model and Problem Statement

We consider a wireless link where the transmitter can transmit at \( n \) possible transmission rates: \( r_1, r_2, \ldots, r_n \). Let the set of these \( n \) transmission rates be denoted by \( \mathcal{R} \). Without loss of generality, we assume that \( r_1 < r_2 < \ldots < r_n \). Corresponding to each transmission rate \( r_i \), there is a rate admissibility probability \( \theta^*_i \) which denotes the probability with which the transmission will be successful at rate \( r_i \), i.e., \( \mathbb{P}\{\text{transmission at rate } r_i \text{ goes through}\} = \theta^*_i \). Let \( \theta^* = (\theta^*_1, \theta^*_2, \ldots, \theta^*_n) \). The probability of success for lower transmission rates is higher, i.e., we have \( 1 = \theta^*_1 > \theta^*_2 > \ldots > \theta^*_n \). The assumption that transmission at the lowest rate is always successful is without loss of generality since we can always let \( r_1 = 0 \).

We elaborate on the above model further by looking at the wireless channel in more detail. Consider a random channel \( (h(t))_{t \geq 0} \) which can be in one of the following \( n \) states (at any time \( t \)): \( h_1, h_2, \ldots, h_n \). Let \( \mathcal{H} = \{h_1, h_2, \ldots, h_n\} \). Let the corresponding probabilities associated with these channel states be \( \nu^* = (\nu^*_1, \nu^*_2, \ldots, \nu^*_n) \), i.e., \( \mathbb{P}\{h(t) = h_i\} = \nu^*_i \). At each time slot \( t \), the channel state \( h(t) \) is drawn independently from the above distribution. Each channel state admits a maximum possible transmission rate, i.e., corresponding to each channel state \( h_i \in \mathcal{H} \), we have a maximum possible rate \( r_i \) which can be successfully transmitted. Without loss of generality, we assume that \( h_1, h_2, \ldots, h_n \) are ordered in the increasing order of their respective maximum admissible transmission rates, i.e., \( r_1 < r_2 < \ldots < r_n \). As before, let \( \mathcal{R} = \{r_1, r_2, \ldots, r_n\} \). Note that if the channel is in state \( h_k \), it can admit transmission rates \( r_i, 1 \leq i \leq k \). Therefore, for any rate \( r_i \), the probability of being successfully transmitted at any time \( t \) is \( \sum_{j=i}^{n} \nu^*_j \). From the definition of \( \theta^*_i \), we have \( \theta^*_i = \sum_{j=i}^{n} \nu^*_j \).

Our goal is to use the communication channel as efficiently as possible. Hence, the aim is to transmit at the optimal transmission rate, i.e., the transmission rate that maximizes the expected throughput at each time slot. If the channel state probabilities or the rate admissibility probabilities are known, this essentially translates to solving the following optimization problem to find the optimal rate \( r^* \):

\[
r^* = \arg\max_{r_i \in \mathcal{R}} r_i \times \sum_{j=i}^{n} \nu^*_j \equiv \arg\max_{r_i \in \mathcal{R}} r_i \times \theta^*_i \tag{1}
\]

The challenge is that the channel state probabilities or the rate admissibility probabilities are unknown. Therefore, we cannot solve the optimization problem (1) exactly. Our aim is to design an algorithm that determines the rate of transmission at each time slot such that our expected throughput over a large time-horizon is as close to the optimal expected throughput as possible.

We call the maximization problem in (1), the rate selection problem where we adapt the channel transmission rate to the unknown success probabilities \( \theta^*_i \), which have to be learned either directly or indirectly through some learning algorithm. The rate selection problem has similarities to the multi-armed bandit problem. Each transmission rate can be treated as a possible arm to pull in a multi-armed bandit scenario. The aim is to transmit at the optimal rate (pulling the optimal arm) at each time slot to minimize the expected regret. The major difference between our problem setup and the multi-armed bandit problem is the fact that the rate admissibility probabilities for different rates (components of \( \theta^* \)) are correlated and not independent of each other. This difference gives rise to difficulties and challenges which do not arise in the traditional multi-armed bandit framework.

We now set the notation for the rest of the paper. Let the transmission rate at each time slot \( t \) be denoted by \( r(t) \), which belongs to the set \( \{r_1, r_2, \ldots, r_n\} \). Also, let the channel state at time \( t \) be \( h(t) \), where \( h(t) = h_j \), for some \( j \in \{1, 2, \ldots, n\} \). At each time slot \( t \), we observe a random variable \( X(t) = f(h(t), r(t)) \equiv I\{r(t) \leq r_j\} \), i.e., the random variable \( X(t) \) is 1 if the transmission at rate \( r(t) \) was successful and 0 otherwise. Let \( X(t) \in \mathcal{X} \), where \( \mathcal{X} \equiv \{0, 1\} \). If the rate at which we transmit is less than or equal to the maximum admissible rate of the channel state then the throughput is equal to the transmission rate, otherwise the throughput is 0. The optimization problem (1) can then be rewritten as:

\[
r^* = \arg\max_{r_i \in \mathcal{R}} E[r(t) \times X(t)|r(t) = r_i, \theta^*] \tag{2}
\]
For ease of exposition, let \( i^* \) denote the index corresponding to the optimal rate, i.e., \( r^* = r_{i^*} \). Let the probability distribution for the random transmission outcome \( X(t) = f(h(t), r) \) at each time slot \( t \) (given the transmission rate \( r \) and the underlying rate admissibility distribution parameter \( \theta \)) be represented by \( p(x; r, \theta) \). Note that \( p(x; r, \theta) \) is a Bernoulli distribution as \( X(t) \in \{0, 1\} \). For any parameter \( \theta \), the optimal transmission rate is given by \( r_{opt}(\theta) = \arg \max_{r \in \mathbb{R}} \mathbb{E}[r(t)X(t)|r(t) = r, \theta] \). Let \( r^* = r_{i^*} = r_{opt}(\theta^*) \). Since we do not know the true parameter \( \theta^* \), we need to design an algorithm that minimizes the number of times we transmit at sub-optimal rates, i.e., the number of times we select sub-optimal actions. We define the (expected) regret/loss for \( t \) as
\[
\Delta_t = r_{i^*}(t) - r_{opt}(\theta^*). \tag{5}
\]

### III. ALGORITHMS

In this section, we first briefly discuss the Thompson Sampling (TS) algorithm (see [20], [14], [15]). Although the Thompson sampling algorithm for the standard multi-armed bandits problem with Bernoulli rewards does not apply directly to our problem, we build on it to design MTS, a Modified Thompson Sampling algorithm. However, a more general version of the Thompson sampling algorithm (see [15], Algorithm 1) applies to our problem but is not feasible. We will illustrate why this general TS algorithm is not suitable for our problem. We then present our algorithm which is inspired by the TS algorithm for Bernoulli bandits (see [20], [14]) and is referred to as MTS (see Algorithm 2). In subsequent sections, we will present theoretical guarantees on the performance of MTS. We also provide simulation results to corroborate the theoretical claims.

#### A. Thompson sampling algorithm

In the standard stochastic multi-armed bandit problem, we have several actions (or arms) available to us and at every time slot, we need to choose one of the available actions to play. Once an action is played, we receive a random reward. Corresponding to every action, the random reward is drawn from a probability distribution with a finite expected value. The reward for the action played is independent and identically distributed (i.i.d.) at every time slot.

The objective of the problem is to design an algorithm that determines the best action to play at any time slot, i.e., the action with the maximum expected value of the reward outcome. The algorithm has access to the history of actions played and the reward outcomes until the latest time slot and can use this history to choose the next action. The multi-armed bandit problem is a well-studied problem in literature (see [21] for a survey).

Thompson sampling is a popular algorithm that is applied to solve the multi-armed bandit problem. In [14], Agrawal and Goyal obtain an upper bound on the regret (expected reward loss because of the non-optimal actions played) due to Thompson sampling for Bernoulli as well as non-Bernoulli rewards, and show that it matches a lower bound due to Lai and Robbins (see [16]) in the asymptotic regime (when the number of times the bandit is played approaches infinity).

Thompson sampling can also be used in settings more general than the multi-armed bandit setting (for example [15]). While there are no known lower bounds in all such cases, it has been shown in [15] that the regret is still upper bounded logarithmically as a function of time \( T \). The optimal link rate selection problem falls in the more general problem setup considered in [15]. Therefore, in principle, one can use the general Thompson sampling algorithm (Algorithm 1) for the problem we consider. However, a direct implementation is infeasible as we discuss next.

#### Algorithm 1 General Thompson sampling

**initialize prior \( p_0(1) \) (for channel state probability vector \( \nu \)).**

**for each** \( t = 1, 2, \ldots \):

1. Draw \( \nu(t) \sim p_\nu(t) \). Compute \( [\theta(t)]_i = \sum_{j=1}^n [\nu(t)]_j \).
2. Transmit at rate \( r_{i(t)} \), where \( r_{i(t)} = r_{opt}(\theta(t)) \).
3. Observe the random transmission outcome \( X(t) \).
4. (Prior Update) Set \( p_\nu(t+1) \propto f(X(t)|\nu)p_\nu(t) \).

**end for**

#### B. Challenges

Following are the major challenges which arise if we use Algorithm 1 for our problem:

1. While dealing with the rate admissibility probabilities \( \theta \), it is difficult to come up with a feasible prior distribution \( p_\nu(t) \) for running the general Thompson sampling algorithm. Since the rate admissibility probability distribution is not multinomial and has interdependent components, the prior required would be complicated and difficult to update. However, one can use Thompson sampling to estimate the channel state probability \( \nu \) (Algorithm 1), but it comes at a huge computational cost as we discuss next.

2. If we deal with the multinomial channel state distribution \( \nu \), we can use the popular Dirichlet distribution as the prior over \( \mathcal{V} \). But since we observe only the outcome of our transmission and not the exact channel state, the posterior update for the Dirichlet prior distribution may require exponentially increasing storage and computational power depending on the trajectory of the algorithm. For example, let us consider the case where \( n = 3 \), i.e., there are 3 possible states the channel can take. At \( t = 1 \), we start with a Dirichlet distribution as prior with parameters \( (1, 1, 1) \), i.e., \( Dir(1, 1, 1) \). Suppose at \( t = 1 \), we transmit at rate \( r_2 \) and it is successful. We simply know that the channel is either in channel state 2 or 3. Therefore, after standard calculations, the prior becomes:

\[
\frac{B((1, 2, 1))}{B((1, 2, 1)) + B((1, 1, 2))} Dir(1, 2, 1) + \]
Algorithm 2 Modified Thompson sampling algorithm
for each rate $r_i, i = 1, 2, ..., n$, set $S_i = 0$ and $F_i = 0$.
for each $t = 1, 2, ...$
1) For all rates $r_i$, draw $\theta_i(t) \sim \text{Beta}(S_i + 1, F_i + 1)$.\footnote{Beta($a, b$) refers to the beta distribution whose probability density function is given by $p_{a,b}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, x \in [0,1]$, where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.}
2) Transmit at rate $r_{i(t)}$, where $i(t) = \arg\max_i r_i \theta_i(t)$.
3) Observe the random transmission outcome $X(t)$.
4) (Posterior Update for Prior) 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

\[ \frac{B((1,1,2))}{B((1,1,2))+B((1,1,2))} \text{Dir}(1,1,2) \{ \text{where } B(\alpha) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^k \alpha_i)} \text{ and } \alpha = (\alpha_1, \alpha_2, ..., \alpha_k) \} \]

Clearly, we now need to store 2 sets of Dirichlet parameters instead of 1. As the number of iterations increase, the number of parameters to be stored and evaluated increases exponentially. After rate slot times, the number of Dirichlet distribution parameters to be stored and evaluated could be as high as $2^t$. This renders the algorithm infeasible due to memory and computational constraints.

\[ \text{Algorithm } \text{Modified Thompson Sampling algorithm} \]

Although it is difficult to find a prior for $\theta$ in the general Thompson sampling algorithm, we would still like to work with $\theta$ instead of $\nu$ as the limited feedback that we get from the system does not give us exact CSI. The only information we get is whether transmission at a certain rate was successful or not. Hence, intuitively, it makes more sense to work with $\theta$ instead of $\nu$.

Therefore, in MTS (Algorithm 2), since it is not possible to have one prior for the vector $\theta$, we maintain $n-1$ priors for the scalar components of $\theta$, i.e., $\theta_2, ..., \theta_n$. Note that $\theta_1 = 1$ for all $\theta$, so we only need $n-1$ priors. This decoupling allows us to use the simple beta prior for the components of $\theta$. At each iteration we only update the prior of the component for which the rate at which we transmit provides conclusive information. As we shall establish in the sequel, this decoupling yields a computationally light solution that achieves a logarithmic (or smaller regret) as a function of time $T$. Note that it is a bit surprising that one is still able to obtain logarithmic or lower regret even though the estimate $\hat{\theta}(t) = (\hat{\theta}_1(t), \hat{\theta}_2(t), ..., \hat{\theta}_n(t))$ (stochastic estimate of $\theta^*$) in Algorithm 2 does not conform to the condition $\theta_1(t) > \theta_2(t) > ... \theta_n(t)$ imposed by the true model $\theta^*$.

IV. PERFORMANCE ANALYSIS: An Upper Bound

To study MTS, we cannot directly use Agrawal and Goyal’s analysis (see [14]). Instead we modify their analysis to show that our algorithm achieves logarithmic or constant regret (depending upon the problem parameters). For our analysis, we adopt the definitions and notation from [14], which we reproduce here for convenience.

Definition 1. (Parameters $N_i(t), i(t), S_i(t)$ and $\mu_i(t)$). Let $r_{i(t)}$ denote the transmitted rate at time $t$, where $i(t)$ denotes the index of the rate in the set $R$. Let $N_i(t)$ denote the number of times rate $r_i$ has been transmitted until time $t - 1$. Let $S_i(t)$ denote the number of successful transmissions of the rate $r_i$ until time $t - 1$. Moreover, $\mu_i(t)$ is defined as the empirical mean of the transmission outcomes for a rate $r_i$ until time $t - 1$, i.e., $\mu_i(t) = \frac{\sum_{t=1}^{t-1} S_i(t)}{N_i(t)}$.

To analyze the performance of MTS theoretically, we will first upper bound the number of times we transmit at any sub-optimal rate $r_i (i \neq i^*)$ until time $T$. Eventually, to obtain the upper bound on total regret, we will simply sum the regret until time $T$ due to each sub-optimal rate of transmission.

Definition 2. (Thresholds $x_i, y_i$). For each rate $r_i (i \neq i^*)$, we will choose two thresholds $x_i$ and $y_i$ such that $r_i \theta_i^* > r_i x_i < r_i y_i < r_i \theta_i^*$. The choice of exact values of $x_i$ and $y_i$ will be presented in the proof.

Definition 3. (Events $E_i^a(t), E_i^b(t)$). We define the event $E_i^a(t)$ as the event such that $\mu_i(t) \leq x_i$. Similarly, $E_i^b(t)$ is the event such that $\theta_i(t) \leq y_i$.

$E_i^a(t)$ defines the event that the empirical average of the outcomes of transmission at rate $r_i$ (until time $t - 1$) does not deviate too much from the true expected value $\theta_i^*$. Similarly, $E_i^b(t)$ defines the event that the sampled parameter for the rate $r_i$ (by MTS at time $t$) does not deviate too much from $\theta_i^*$.

Definition 4. (Filtration $\mathcal{F}_{t-1}$). We define the filtration $\mathcal{F}_{t-1}$ as the history of rates transmitted and their outcomes until time $t - 1$, i.e., $\mathcal{F}_{t-1} = \{i(j), X(j); j = 1, ..., t - 1\}$.

Definition 5. (Parameters $\tau_i$ and $p_{i,t}$). Let $\tau_i$ denote the time when the optimal rate $r_i^*$ is transmitted the $i$th time (for $i \geq 1$). Also, let $\tau_0 = 0$. We define the probability $p_{i,t}$ as, $p_{i,t} = P(t, \theta_i(t) > r_i y_i | \mathcal{F}_{t-1}) = P(\theta_i(t) > \frac{y_i}{r_i} | \mathcal{F}_{t-1})$.

A point worth noting is that, for every rate $r_i$, $\mathcal{F}_{t-1}$ determines $p_{i,t}, S_i(t), N_i(t), \mu_i(t)$, the distribution of $\theta_i(t)$ and whether the event $E_i^b(t)$ is true or not. To bound the expected number of times we transmit at rate $r_i$, as in [14], we split the expectation into three different terms based on the occurrences of the events $E_i^a(t)$ and $E_i^b(t)$:

\[ \mathbb{E}[N_i(T + 1)] = \sum_{t=1}^{T} \mathbb{P}(i(t) = i) \]
\[ = \sum_{i=1}^{T} \mathbb{P}(i(t) = i, E_i^a(t), E_i^b(t)) + \sum_{i=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^a(t)}, E_i^b(t)) \]
\[ + \sum_{i=1}^{T} \mathbb{P}(i(t) = i, E_i^a(t), \overline{E_i^b(t)}) \]

(3)

where $\overline{A}$ denotes the complement of event $A$.

Remark: To upper bound the LHS above, we will find upper bounds for the three terms on RHS separately and subsequently add them.
We start with analyzing the first term, i.e., \( \sum_{t=1}^{T} \mathbb{P}(i(t) = i, E^o_t(t), E^i_t(t)) \). We obtain a lemma (as in [14]) which establishes a relationship between the probabilities of choosing a sub-optimal rate \( r_i \) and the probability of choosing the optimal rate \( r^*_i \) (given the filtration \( \mathcal{F}_{t-1} \), along with the occurrence of events \( E^o_t(t), E^i_t(t) \)) in terms of \( p_{i,t} \).

**Lemma 1.** For all \( t \in [1, T] \), and \( i \neq \star \), we have:

\[
\mathbb{P}(i(t) = i, E^o_t(t), E^i_t(t)|\mathcal{F}_{t-1}) \leq \left(1 - \frac{p_{i,t}}{p_{i,t}}\right) \mathbb{P}(i(t) = i^*, E^o_t(t), E^i_t(t)|\mathcal{F}_{t-1}).
\]

**Proof.** Since \( \mathcal{F}_{t-1} \) determines the status of the event \( E^i_t(t) \), we assume that the event took place as otherwise the LHS of the result is 0 and hence the lemma holds trivially. Therefore, we just need to show the following:

\[
\mathbb{P}(i(t) = i|\mathcal{F}_{t-1}, E^o_t(t)) \leq \left(1 - \frac{p_{i,t}}{p_{i,t}}\right) \mathbb{P}(i(t) = i^*|\mathcal{F}_{t-1}, E^i_t(t)).
\]

For any sub-optimal rate of transmission \( r_i \), i.e., \( i \neq \star \), we have:

\[
\mathbb{P}(i(t) = i|\mathcal{F}_{t-1}, E^o_t(t)) \\
\leq \mathbb{P}(r_j \theta_j(t) \leq r_i | \mathcal{F}_{t-1}, E^o_t(t)) \\
= \mathbb{P}(r_j \theta_j(t) \leq r_i | \mathcal{F}_{t-1}) \times \mathbb{P}(r_j \theta_j(t) \leq r_i, \forall j \neq i^* | \mathcal{F}_{t-1}, E^o_t(t)) \\
= \left(1 - \frac{p_{i,t}}{p_{i,t}}\right) \mathbb{P}(i(t) = i^*|\mathcal{F}_{t-1}, E^i_t(t)).
\]

The first inequality above follows from the fact that the event \( \{i(t) = i|E^o_t(t)\} \) is a subset of the event \( \{r_j \theta_j(t), \forall j \leq r_i | E^o_t(t)\} \). Also, the first equality follows from the fact that the beta priors for different rates at any time \( t \) are independent of each other given the filtration \( \mathcal{F}_{t-1} \). Conditioning on the event \( E^i_t(t) \) retains the independence between \( \theta_i(t) \) and \( \theta_j(t), \forall j \neq i^* \). Similarly, we have:

\[
\mathbb{P}(i(t) = i^*|\mathcal{F}_{t-1}, E^o_t(t)) \\
\geq \mathbb{P}(r_i \theta_i(t) > r_j \theta_j(t), \forall j \neq i^* | \mathcal{F}_{t-1}, E^o_t(t)) \\
= \mathbb{P}(r_i \theta_i(t) > r_j \theta_j(t) | \mathcal{F}_{t-1}) \times \mathbb{P}(r_j \theta_j(t) \leq r_i, \forall j \neq i^* | \mathcal{F}_{t-1}, E^o_t(t)) \\
= p_{i,t} \times \mathbb{P}(r_j \theta_j(t) \leq r_i, \forall j \neq i^* | \mathcal{F}_{t-1}, E^o_t(t)).
\]

Combining the above two inequalities, we get (4) and hence the lemma. \( \square \)

Using Lemma 1 and the analysis preceding Lemma 2 in [14], we get:

\[
\sum_{t=1}^{T} \mathbb{P}(i(t) = i, E^o_t(t), E^i_t(t)) = \sum_{j=0}^{T-1} \mathbb{E}\left[\frac{1}{p_{i,t}} - 1\right].
\]

We can upper bound the term \( \mathbb{E}\left[\frac{1}{p_{i,t}} - 1\right] \) in the above equation using Lemma 2 in Agrawal and Goyal’s paper (see [14]) by replacing \( y_i \) in their lemma with \( \frac{r_i \theta_i}{r_i} \). Hence, combining (5) with Lemma 2 in from [14]:

\[
\sum_{t=1}^{T} \mathbb{P}(i(t) = i, E^o_t(t), E^i_t(t)) \\
\leq \frac{24}{\Delta^2} + \sum_{j=0}^{T-1} \Theta(e^{-\Delta^2 j} + e^{-D_{i,j}} + \frac{1}{(k+1)\Delta^2} - 1).
\]

We obtain a lemma (as in [14]) which determines the status of the event \( 1 \) in terms of \( \mathcal{F}_{t-1} \) in the above equation. Conditioning on \( \mathcal{F}_{t-1} \), we have:

\[
\mathbb{P}(i(t) = i, E^o_t(t), E^i_t(t)) = 0\]
Combining (7) and Lemma 2 with the fact that \( L_i(T) = \frac{\log T}{D(x_i, c_i)} = (1 + \epsilon)^2 \frac{\log T}{D(\theta_i^*, r_i^*, c_i)} \):

\[
\sum_{t=1}^{T} P(i(t) = i, E_t^i(t), E_t^i(t)) \leq (1 + \epsilon)(1 + \epsilon^2) \frac{\log T}{D(\theta_i^*, r_i^*, c_i)}
\]

We are now left with the third and the final term in (3). Using Lemma 3 in Agrawal and Goyal’s paper (see [14]) we get:

\[
\sum_{t=1}^{T} P(i(t) = i, E_t^i(t)) \leq \frac{1}{D(x_i, \theta_i^*)} + 1
\]

Moreover, using the fact that \( D(x_i, \theta_i^*) = D(\theta_i^*, r_i^*, c_i) \), after some manipulations, we can get:

\[
x_i - \theta_i^* \geq \epsilon \times \frac{\log \left( \frac{D(\theta_i^*, r_i^*, c_i)}{x_i - \theta_i^*} \right)}{1 + \epsilon}
\]

Above inequality gives \( \frac{1}{D(x_i, \theta_i^*)} \leq \frac{1}{2(x_i - \theta_i^*)^2} = O(\frac{1}{\epsilon^2}) \). Using the above fact in (9):

\[
\sum_{t=1}^{T} P(i(t) = i, E_t^i(t)) \leq O(\frac{1}{\epsilon^2})
\]

Combining (6), (8) and (10), we get:

\[
E[N_i(T + 1)] \leq O(1) + \frac{I(r_i^*, \theta_i^* \leq r_i^*)}{D(\theta_i^*, r_i^*, c_i)} + O\left(\frac{n}{\epsilon^2}\right)
\]

\[
\leq (1 + \epsilon') \frac{I(r_i^*, \theta_i^* \leq r_i^*)}{D(\theta_i^*, r_i^*, c_i)} + O\left(\frac{n}{\epsilon^2}\right)
\]

where \( \epsilon' = 3\epsilon \). Therefore, from (11), we get the following theorem:

**Theorem 1.** For the \( n \)-rates optimal link rate selection problem, MTS algorithm has the following expected regret until time \( T \):

\[
E[l(T)] \leq (1 + \epsilon') \frac{I(r_i^*, \theta_i^* \leq r_i^*)}{D(\theta_i^*, r_i^*, c_i)} \Delta_i + O\left(\frac{n}{\epsilon^2}\right)
\]

for any \( \epsilon \in (0, 1) \), where \( \Delta_i = r_i^* - r_i^* \).

**A. Discussion**

The idea of decoupling the components of \( \theta \) and using separate priors can be used to design another algorithm for the optimal link rate selection problem which we present as Algorithm 3. This algorithm combines the idea of decoupling components of \( \theta \) with the Thompson sampling algorithm for non-Bernoulli bandits presented in [14]. For Algorithm 3, the following result is an immediate consequence of Theorem 1 in [14]:

**Algorithm 3** Algorithm motivated by prior work in [14]

for each rate \( r_i, i = 1, 2, \ldots, n \), set \( S_i = 0 \) and \( F_i = 0 \).

for each \( t = 1, 2, \ldots \):

1. For all rates \( r_i \), draw \( \mu_i(t) \sim \text{Beta}(S_i + 1, F_i + 1) \).
2. Transmit at rate \( r_i(t) \), where \( i(t) = \arg\max_i \mu_i(t) \).
3. Observe the normalized random transmission throughput \( Y(t) = \frac{r_i(t)}{r_n^*} X(t) \). Draw temp \~ Bernoulli(Y(t)).
4. (Posterior Update for Prior) If temp = 1, set \( S_{i(t)} = S_{i(t)} + 1 \). Else if temp = 0, set \( F_{i(t)} = F_{i(t)} + 1 \).

**end for**

**Theorem 2.** For the \( n \)-rates optimal link rate selection problem, Algorithm 3 has the following expected regret until time \( T \):

\[
E[l(T)] \leq (1 + \epsilon') \frac{I(r_i^*, \theta_i^* \leq r_i^*)}{D(\theta_i^*, r_i^*, c_i)} \Delta_i + O\left(\frac{n}{\epsilon^2}\right)
\]

for any \( \epsilon \in (0, 1) \), where \( \Delta_i = r_i^* - r_i^* \).

One of the contributions of this paper is to show that the decoding of transmission rates in our proposed MTS algorithm is superior to Algorithm 3. To see this, note that MTS has \( O(1) \) regret for certain problem parameters (Case 2) whereas Algorithm 3 can only be proven to have \( O(\log T) \) regret regardless of problem parameters. Additionally, the constant factor associated with the logarithmic regret term for MTS is \( 1/\Delta_i \), whereas Algorithm 3 has a constant factor of \( 1/\Delta_i \), \( D(\theta_i^*, r_i^*, c_i) \) will be less than \( D(\theta_i^*, r_i^*, c_i) \) since the multiplication by \( \Delta_i \) in the former case will drive the two Bernoulli distributions closer, effectively reducing the KL-divergence between them. Simulation results also confirm these findings.

**V. PERFORMANCE ANALYSIS: A LOWER Bound**

In this section, we prove a lower bound for a special case of the optimal link rate selection with 3 channel states and show that MTS is optimal in this case, i.e., the constant factor associated with the logarithmic regret term in MTS is tight. We will use Lai and Robbins style of analysis to obtain the lower bound (see [16] for details). Recall that, for the optimal link rate selection problem with three channel states, we have: \( \mathcal{H} = \{h_1, h_2, h_3\} \), \( \mathcal{R} = \{r_1, r_2, r_3\} \) with \( r_1 < r_2 < r_3 \). Also, the channel state probability vector is given by \( \nu = (\nu_1, \nu_2, \nu_3) \). The rate admissibility probability vector is given by \( \theta = (\theta_1^*, \theta_2^*, \theta_3^*) \), where \( \theta_i^* = \sum_{r_j \leq r_i} \nu_j \). Typically, the lowest rate of transmission is 0, so we assume \( r_1 = 0 \).

Since \( r_1 \) is zero, we will only consider the cases where either rate \( r_2 \) or rate \( r_3 \) is optimal.

**Case 1:** \( r_2 \) is optimal.

Let us start with the case where the rate \( r_2 \) is the unique optimal rate, i.e., \( r_2 \theta_2^* > r_3 \theta_3^* \). Consider \( \theta' = (\theta_1^* = 1, \theta_2^*, \theta_3^*) \), such that \( r_3 \theta_3^* > r_2 \theta_2^* \), i.e., \( \theta' \) is such that the unique optimal transmission rate for \( \theta' \) is \( r_3 \) and the first two components of \( \theta' \) and \( \theta^* \) are the same.
We will choose an appropriate value of $c_T$ later. Considering $L(\mathcal{F}_T) \leq c_T$ first:

$$\mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) = \sum_{\mathcal{F}_T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T} e^{-L(\mathcal{F}_T)} \mathbb{P}_{\theta^*}(\mathcal{F}_T) \geq e^{-c_T} \sum_{\mathcal{F}_T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T} \mathbb{P}_{\theta^*}(\mathcal{F}_T) = e^{-c_T} \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T).$$

We can rewrite the above inequality as:

$$\mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) \leq e^{c_T} \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) \tag{13}$$

From the law of total probability, we have:

$$\mathbb{P}_{\theta^*}(N_3(T + 1) = n) = \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) + \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) > c_T) \leq e^{c_T} \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) + \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) > c_T) \leq e^{c_T} \mathbb{P}_{\theta^*}(N_3(T + 1) = n) + \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) > c_T) \tag{14}$$

where the second last inequality follows from (13) and the last inequality follows from the fact that $\{\mathcal{F}_T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T\} \subseteq \{\mathcal{F}_T : N_3(T + 1) = n\}$. As mentioned previously, we need to show that probability (under $\theta^*$) of $N_3(T + 1)$ being less than a certain time-dependent threshold $(f_T)$ approaches 0 as time approaches $\infty$.

$$\mathbb{P}_{\theta^*}(N_3(T + 1) \leq f_T) = \sum_{n \leq f_T} \mathbb{P}_{\theta^*}(N_3(T + 1) = n) \leq e^{c_T} \sum_{n \leq f_T} \mathbb{P}_{\theta^*}(N_3(T + 1) = n) \leq e^{c_T} \sum_{n \leq f_T} \mathbb{P}_{\theta^*}(N_3(T + 1) = n, L(\mathcal{F}_T) > c_T) \leq e^{c_T} \mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) > c_T) \tag{15}$$

Considering the first term on the RHS and using Markov’s inequality, we get:

$$\mathbb{P}_{\theta^*}(T : N_3(T + 1) = n, L(\mathcal{F}_T) \leq c_T) \leq \frac{\mathbb{E}_{\theta^*}(T - N_3(T + 1))}{T - f_T} \tag{16}$$

Under $\theta^*$, $r_3$ is the optimal transmission rate, therefore $T - N_3(T + 1)$ is the number of times the policy transmits at a sub-optimal rate. We want this to be small, hence we choose $\mathbb{E}_{\theta^*}(T - N_3(T + 1)) = o(T^\alpha)$, for some $\alpha \in (0, 1)$. Using (15):

$$\mathbb{P}_{\theta^*}(T : N_3(T + 1) \leq f_T) = o(T^{\alpha - 1}) \tag{16}$$
For ease of notation, let $Y(s) = \sum_{j=1}^{s} \{ (X_{3}(j) \log(\frac{\theta_{3}^{*}}{\theta_{3}}) + (1 - X_{3}(j)) \log(\frac{1 - \theta_{3}^{*}}{1 - \theta_{3}}) \}$. Now, we consider the second term on the RHS of (14):

$$
P_{\theta_{3}}(N_{3}(T + 1) \leq f_{T}, L(\mathcal{F}_{T}) > c_{T})$$

$$= \sum_{s=1}^{f_{T}} P_{\theta_{3}}(N_{3}(T + 1) = s, L(\mathcal{F}_{T}) > c_{T})$$

$$= \sum_{s=1}^{f_{T}} P_{\theta_{3}}(N_{3}(T + 1) = s, Y(s) > c_{T})$$

$$\leq \sum_{s=1}^{f_{T}} P_{\theta_{3}}(N_{3}(T + 1) = s, \max_{s \in 1,2,\ldots,f_{T}} Y(s) > c_{T})$$

$$= \max_{s \in 1,2,\ldots,f_{T}} Y(s) > c_{T})$$

where the first inequality follows from the fact that the event \( \{ Y(s) > c_{T} \} \subset \{ \max_{s \in 1,2,\ldots,f_{T}} Y(s) > c_{T} \} \), \( 1 \leq s \leq f_{T} \). The last step follows from the fact that \( P(A,B) \leq P(A) \). Now, by Strong Law of Large Numbers, we have

$$\lim_{T \to \infty} f_{T} = \lim_{T \to \infty} \frac{1}{T} \sum_{s=1}^{\min(s, f_{T})} X_{3s} \log\left(\frac{\theta_{3}}{\theta_{3}^{*}}\right) + (1 - X_{3s}) \log\left(\frac{1 - \theta_{3}}{1 - \theta_{3}^{*}}\right)$$

almost surely. It is also easy to show that if \( X_{t} \to C \) a.s., then \( \max_{t} X_{t} \to C \) almost surely. Therefore, if we choose \( \frac{c_{T}}{f_{T}} > D(\theta_{3}^{*}|\theta_{3}) \), then \( P_{\theta_{3}}(N_{3r} \leq f_{T}, L(\mathcal{F}_{T}) > c_{T}) \to 0 \) as \( f_{T} \to \infty \) almost surely. This takes care of the second term on the RHS of (14).

Combining (14) and (16), we observe that we need

$$e^{c_{T}o(T^{\alpha} - 1)} \to 0 \text{ as } T \to \infty \text{ so that } P_{\theta_{3}}(N_{3}(T + 1) \leq f_{T}) = 0 \text{ as } T \to 0 \text{.}$$

Therefore:

$$e^{c_{T}o(T^{\alpha} - 1)} = o(e^{(\alpha - 1)\log T + c_{T}})$$

Thus, we need \((\alpha - 1)\log T + c_{T} \to \infty \text{ as } T \to \infty \). This is true if we choose \( c_{T} = \frac{1 - \alpha}{1 + \gamma} \log T \), where \( \gamma > 0 \). Also, we choose \( f_{T} = \frac{1 - \delta}{\log\left(\frac{\theta_{3}}{\theta_{3}^{*}}\right)} \), \( \delta \in (0, 1) \). These choices of \( f_{T}, c_{T} \) satisfy the requirements that \( f_{T} \to \infty \text{ as } T \to \infty \) and that \( \frac{c_{T}}{f_{T}} > D(\theta_{3}^{*}|\theta_{3}) \). Let \( \rho(T) = P_{\theta_{3}}(N_{3}(T + 1) \leq f_{T}) \to 0 \text{ as } T \to \infty \text{.}$$

Therefore, we conclude that:

$$\lim_{T \to \infty} \rho(T) = 0 \quad (17)$$

(17) is true for any \( \delta \in (0, 1), \alpha \in (0, 1), \gamma > 0 \) and any policy that transmits at a sub-optimal rate for \( o(T^{\alpha}) \) times on average. Using Markov’s inequality, we get:

$$D(\theta_{3}^{*}|\theta_{3}) \text{E}_{\theta_{3}}[N_{3}(T + 1)] \geq 1 - \rho(T)$$

Since, the above equation is true for any \( \delta \in (0, 1) \) and \( \alpha \in (0, 1) \), taking limits on both sides, we get:

$$\lim_{T \to \infty} \frac{\text{E}_{\theta_{3}}[N_{3}(T + 1)]}{\log T} \geq \frac{D(\theta_{3}^{*}|\theta_{3})}{1 - \delta} \quad (18)$$

Only thing left for us to do now is to choose an appropriate \( \theta_{3}^{*} \). Note that we want \( r_{2}\theta_{2} < r_{3}\theta_{3}^{*} \), hence we can choose any \( \theta_{3}^{*} \) such that \( \theta_{3}^{*} = \min\{ \frac{r_{2}}{r_{3}} \theta_{2} + \epsilon, \theta_{2} \}, \epsilon > 0 \). Using this fact in (18), we get:

$$\lim_{T \to \infty} \frac{\text{E}_{\theta_{3}}[N_{3}(T + 1)]}{\log T} \geq \frac{1}{D(\theta_{3}^{*}|\theta_{3})}$$

We now consider the case when \( r_{3} \) is the optimal rate.

**Case 2:** \( r_{3} \) is optimal.

In the case when \( r_{3} \) is optimal, if \( \frac{r_{2}\theta_{2}}{r_{3}} > 1 \), MTS achieves \( O(1) \) regret and hence we can use the trivial lower bound of \( 0 \). On the other hand, if \( \frac{r_{2}\theta_{2}}{r_{3}} \leq 1 \), we can choose a \( \theta_{3}^{*} \) such that \( r_{2}\theta_{2} > r_{3}\theta_{3}^{*} \). The same analysis as that of Case 1 would then hold. Note that \( \frac{r_{2}\theta_{2}}{r_{3}} \leq 1 \) is always true since \( r_{2} > r_{3} \) and \( \theta_{2}^{*} \leq 1 \), so Case 1 doesn’t require a trivial lower bound.

Combining Case 1 and Case 2, we get the following theorem:

**Theorem 3.** For the optimal link rate selection problem with three channel states and \( r_{1} = 0 \), the lower bound on expected regret (asymptotically) is given by:

$$\lim_{T \to \infty} \frac{\text{E}_{\theta_{3}}[l(T)]}{\log T} \geq \frac{1}{D(\theta_{3}^{*}|\theta_{3})} \Delta_{i}, i \neq i^{*}$$

where \( \Delta_{i} = r_{i} - r_{i}^{*} - r_{i}^{*} \).

Clearly, the upper bound obtained in Theorem 1 asymptotically matches the lower bound obtained above. A point worth noting here is that although we only obtain the lower bound for the special case of rate selection problem with three channel states, the logarithmic (or smaller) expected regret obtained by MTS in the general case matches the typical state-of-the-art performance achieved by algorithms for the generalizations of the multi-armed bandit problem (see [15]).

**VI. SIMULATION RESULTS**

To corroborate our theoretical results, we implement MTS as well as Algorithm 3 for the optimal link rate selection problem with three channel states. We consider \( r_{1} = 1, r_{2} = 2 \) and \( r_{3} = 3 \). We conduct the following experiments to check the validity of our results:

1) We take \( \nu^{*} = (0.1, 0.1, 0.8) \) (or \( \theta^{*} = (1, 0.9, 0.8) \)) for the first experiment. Under this choice of \( \nu^{*} \), rate \( r_{3} = 3 \) is optimal. Moreover, \( r_{2}\theta_{2} > 1 \) and \( r_{3}\theta_{3} > 1 \). Hence, by Theorem 1, MTS should have \( O(1) \) regret and by Theorem 2, Algorithm 3 should have logarithmic regret.

The results for this experiment are on the left plot in Figure 1. Clearly, the graph confirms the theoretical results. We also repeat the experiment for \( \nu^{*} = (0.3, 0.0, 0.7) \) (or \( \theta^{*} = (1, 0.7, 0.7) \)). This case is also similar to the previous case and the results are plotted on the right graph in Figure 1.

2) We take \( \nu^{*} = (0.3, 0.4, 0.3) \) (or \( \theta^{*} = (1, 0.7, 0.3) \)) for the second experiment. Under this choice of \( \nu^{*} \), rate \( r_{2} = 2 \) is optimal. We have \( r_{2}\theta_{2} > 1 \), but unlike the previous experiment \( r_{3}\theta_{3} \leq 1 \). Hence, by Theorem 1, MTS will have \( O(1) \) regret corresponding to rate \( r_{1} \) and logarithmic regret corresponding to rate \( r_{3} \). On the other hand, by Theorem 2, Algorithm 3 will have logarithmic regret for both \( r_{1} \) and \( r_{3} \). Hence, although
both algorithms will have an overall logarithmic regret, MTS should perform better than Algorithm 3.

The results for this experiment are on the left plot in Figure 2. Clearly, the graph confirms the theoretical results. We also repeat the experiment for \( \nu^* = (0.4, 0.1, 0.5) \) (or \( \theta^* = (1, 0.6, 0.5) \)). This case is also similar to the previous case, although \( r_3 \) is optimal in this case instead of \( r_2 \), and the results are plotted on the right graph in Figure 2.

In all the experiments, MTS outperforms Algorithm 3 by a huge margin as expected.

![Fig. 1. Experiment 1: Implementing MTS and Algorithm 3 for \( \nu^* = (0.1, 0.1, 0.8) \) (left) and \( \nu^* = (0.3, 0.0, 0.7) \) (right). MTS achieves \( O(1) \) regret for both cases while Algorithm 3 achieves logarithmic regret.](image1)

![Fig. 2. Experiment 2: Implementing MTS and Algorithm 3 for \( \nu^* = (0.3, 0.4, 0.3) \) (left) and \( \nu^* = (0.4, 0.1, 0.5) \) (right). Both MTS and Algorithm 3 achieve logarithmic regret but MTS outperforms Algorithm 3 by a huge margin.](image2)

VII. CONCLUSION

In this paper, we consider the optimal link rate selection problem in rapidly varying wireless channels with limited feedback. We propose a low-complexity and low-regret algorithm (MTS) motivated by Thompson sampling to solve the problem. We show that our algorithm MTS achieves logarithmic (or smaller) regret both theoretically as well as experimentally. We also show that for the special case of 3 channel states, the regret achieved by MTS matches the lower bound. Lower bound analysis for the general \( n \)-channel states problem remains open and could be an interesting topic for further research. It will also be interesting to study how the results here can be used to obtain regret bounds for multiple-user models such as the one in [22].

REFERENCES


