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 the feedback about 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 apriori 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]). While many of these works

This research has been supported in part by NSF grants: CNS-NeTS-1717045, CNS-NeTS-1718203, CMMI-SMOR-1562065, CCSS-EARS-1444026, CNS-NeTS-1514127, CNS-ICN-WEN-1719371, and the DTRA grant HDTRA1-15-1-0003. The work of A. Eryilmaz is also supported in part by the QNRF Grant NPRP 7-923-2-344.

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: $r_1, r_2, ..., r_n$. Let the set of these n transmission rates be denoted by $\mathcal{R}$. Without loss of generality, we assume that $r_1 < r_2 < ... < r_n$. Corresponding to each transmission rate r_i, there is a rate admissibility probability θ_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 $1 = \theta_1^* > \theta_2^* > ... > \theta_n^*$. 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): $h_1, h_2, ..., h_n$. Let $\mathcal{H} = \{h_1, h_2, ..., h_n\}$. Let the corresponding probabilities associated with these channel states be $\nu^* = (\nu_1^*, \nu_2^*, ..., \nu_n^*)$, i.e., $P\{h(t) = h_i\} = \nu_i^*, \forall 1 \leq i \leq n, \forall t \geq 0$. 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 $h_1, h_2, ..., h_n$ are ordered in the increasing order of their respective maximum admissible transmission rates, i.e., $r_1 < r_2 < ... < r_n$. As before, let $\mathcal{R} = \{r_1, r_2, ..., r_n\}$. 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 θ_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^*:

$$r^* = \arg\max_{r_i \in \mathcal{R}} \; r_i \times \sum_{j=i}^{n} \nu_j^* \equiv \arg\max_{r_i \in \mathcal{R}} \; r_i \times \theta_i^* \qquad (1)$$

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 θ_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 θ^*) 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 $\{r_1, r_2, ..., r_n\}$. Also, let the channel state at time t be $h(t)$, where $h(t) = h_j$, for some $j \in \{1, 2, ..., n\}$. At each time slot t, we observe a random variable $X(t) = f(h(t), r(t)) \triangleq \mathbb{I}\{r(t) \leq r_j\}$, 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 $\mathcal{X} \triangleq \{0, 1\}$. 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:

$$r^* = \arg\max_{r_i \in \mathcal{R}} \; E[r(t) \times X(t) | r(t) = r_i, \theta^*] \qquad (2)$$

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 θ) 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 θ, the optimal transmission rate is given by $r_{opt}(\theta) = \arg\max_{r \in \mathcal{R}} \mathbb{E}[r(t)X(t)|r(t) = r, \theta]$. Let $r^* = r_{i^*} = r_{opt}(\theta^*)$. Since we do not know the true parameter θ^*, 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 as $\mathbb{E}[l(T)] = \mathbb{E}[\sum_{t=1}^{T} \mathbb{I}\{r(t) \neq r_{i^*}\}(r_{i^*}\theta^*_{i^*} - r(t)\theta^*_{i(t)})]$. Here $i(t)$ denotes the index of $r(t)$, i.e., $r(t) = r_{i(t)}$. The expected regret can also be written as follows:

$$\mathbb{E}[l(T)] = \mathbb{E}[\sum_{i \neq i^*} N_i(T+1)\Delta_i] = \sum_{i \neq i^*} \mathbb{E}[N_i(T+1)]\Delta_i.$$

where, $N_i(T+1)$ is the number of times we transmit at a sub-optimal rate r_i until time T and $\Delta_i = r_{i^*}\theta^*_{i^*} - r_i\theta^*_i$.

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 $p_\nu(1)$ (for channel state probability vector ν).
for each $t = 1, 2, \dots$:

1) Draw $\nu(t) \sim p_\nu(t)$. Compute $[\theta(t)]_i = \sum_{j=i}^{n} [\nu(t)]_j$.
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 $p_\nu(t+1) \propto \mathbb{P}(X(t)|\nu)p_\nu(t)$

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 θ, it is difficult to come up with a feasible prior distribution ($p_\theta(t)$) 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 ν (Algorithm 1), but it comes at a huge computational cost as we discuss next.

2) If we deal with the multinomial channel state distribution ν, 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., $\text{Dir}(1, 1, 1)$. 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 $t = 1, 2, \ldots$:

1) For all rates r_i, draw $\theta_i(t) \sim \text{Beta}(S_i + 1, F_i + 1)$. [1]
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

$\frac{B((1,1,2))}{B((1,2,1))+B((1,1,2))} Dir(1, 1, 2)$ $\{$where $B(\alpha) = \frac{\prod_{i=1}^{k}\Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k}\alpha_i)}$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)\}$. 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 t time slots, 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.

C. Modified Thompson Sampling algorithm

Although it is difficult to find a prior for θ in the general Thompson sampling algorithm, we would still like to work with θ instead of ν 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 θ instead of ν.

Therefore, in MTS (Algorithm 2), since it is not possible to have one prior for the vector θ, we maintain $n-1$ priors for the scalar components of θ, i.e., $\theta_2, ..., \theta_n$. Note that $\theta_1 = 1$ for all θ, so we only need $n-1$ priors. This decoupling allows us to use the simple beta prior for the components of θ. 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 $\theta(t) = (\theta_1(t), \theta_2(t), ..., \theta_n(t))$ (stochastic estimate of θ^*) in Algorithm 2 does not conform to the condition $\theta_1(t) > \theta_2(t) > ... > \theta_n(t)$ imposed by the true model θ^*.

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.

[1] $\text{Beta}(a, b)$ 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 $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Definition 1. *(**Parameters** $N_i(t), i(t), S_i(t)$ and $\hat{\mu}_i(t)$). Let $r_{i(t)}$ denote the transmitted rate at time t, where $i(t)$ denotes the index of the rate in the set $\mathcal{R}$. 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, $\hat{\mu}_i(t)$ is defined as the empirical mean of the transmission outcomes for a rate r_i until time $t-1$, i.e., $\hat{\mu}_i(t) = \frac{\sum_{j=1:i(j)=i}^{t-1} X(j)}{N_i(t)+1}$.*

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** $E_i^{\mu}(t), E_i^{\theta}(t)$) We define the event $E_i^{\mu}(t)$ as the event such that $\hat{\mu}_i(t) \leq x_i$. Similarly, $E_i^{\theta}(t)$ is the event such that $\theta_i(t) \leq y_i$.*

$E_i^{\mu}(t)$ 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 θ_i^*. Similarly, $E_i^{\theta}(t)$ defines the event that the sampled parameter for the rate r_i (by MTS at time t) does not deviate too much from θ_i^*.

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

Definition 5. *(**Parameters** τ_i and $p_{i,t}$). Let τ_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, $p_{i,t} = \mathbb{P}(r_{i^*}\theta_{i^*}(t) > r_i y_i | \mathscr{F}_{t-1}) = \mathbb{P}(\theta_{i^*}(t) > \frac{r_i y_i}{r_{i^*}} | \mathscr{F}_{t-1})$.*

A point worth noting is that, for every rate r_i, $\mathscr{F}_{t-1}$ determines $p_{i,t}$, $S_i(t), N_i(t), \hat{\mu}_i(t)$, the distribution of $\theta_i(t)$ and whether the event $E_i^{\mu}(t)$ 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 $E_i^{\mu}(t)$ and $E_i^{\theta}(t)$:

$$
\begin{aligned}
&\mathbb{E}[N_i(T+1)] \\
&\quad= \sum_{t=1}^{T} \mathbb{P}(i(t) = i) \\
&\quad= \sum_{t=1}^{T} \mathbb{P}(i(t) = i, E_i^{\theta}(t), E_i^{\mu}(t)) \\
&\qquad+ \sum_{t=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^{\theta}(t)}, E_i^{\mu}(t)) + \sum_{t=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^{\mu}(t)})
\end{aligned}
\tag{3}
$$

where $\bar{A}$ 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., $\sum_{t=1}^{T}\mathbb{P}(i(t)=i, E_i^{\theta}(t), E_i^{\mu}(t))$. We obtain a lemma (as in [14]) which establishes a relationship between the probability of choosing a sub-optimal rate r_i and the probability of choosing the optimal rate r_{i^*} (given the filtration $\mathscr{F}_{t-1}$, along with the occurrence of events $E_i^{\theta}(t), E_i^{\mu}(t)$) in terms of $p_{i,t}$:

Lemma 1. *For all $t \in [1,T]$, and $i \neq i^*$, we have:*

$$\begin{aligned}\mathbb{P}(i(t)=i, &E_i^{\mu}(t), E_i^{\theta}(t)|\mathscr{F}_{t-1}) \\ &\leq \frac{(1-p_{i,t})}{p_{i,t}}\mathbb{P}(i(t)=i^*, E_i^{\mu}(t), E_i^{\theta}(t)|\mathscr{F}_{t-1}).\end{aligned}$$

Proof. Since $\mathscr{F}_{t-1}$ determines the status of the event $E_i^{\mu}(t)$, 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:

$$\begin{aligned}\mathbb{P}(i(t)=i|&\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &\leq \frac{(1-p_{i,t})}{p_{i,t}}\mathbb{P}(i(t)=i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)).\end{aligned} \quad (4)$$

For any sub-optimal rate of transmission r_i, i.e., $i \neq i^*$, we have:

$$\begin{aligned}\mathbb{P}(i(t)=i&|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &\leq \mathbb{P}(r_j\theta_j(t) \leq r_i y_i, \forall j|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &= \mathbb{P}(r_{i^*}\theta_{i^*}(t) \leq r_i y_i|\mathscr{F}_{t-1}) \\ &\qquad \times \mathbb{P}(r_j\theta_j(t) \leq r_i y_i, \forall j \neq i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &= (1-p_{i,t}) \times \mathbb{P}(r_j\theta_j(t) \leq r_i y_i, \forall j \neq i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)).\end{aligned}$$

The first inequality above follows from the fact that the event $\{i(t)=i|E_i^{\theta}(t)\}$ is a subset of the event $\{r_j\theta_j(t), \forall j \leq r_i y_i|E_i^{\theta}(t)\}$. 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 $\mathscr{F}_{t-1}$. Conditioning on the event $E_i^{\theta}(t)$ retains the independence between $\theta_{i^*}(t)$ and $\theta_j(t), \forall j \neq i^*$. Similarly, we have:

$$\begin{aligned}\mathbb{P}(i(t)=i^*&|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &\geq \mathbb{P}(r_{i^*}\theta_{i^*}(t) > r_i y_i \geq r_j\theta_j(t), \forall j \neq i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &= \mathbb{P}(r_{i^*}\theta_{i^*}(t) > r_i y_i|\mathscr{F}_{t-1}) \\ &\qquad \times \mathbb{P}(r_j\theta_j(t) \leq r_i y_i, \forall j \neq i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)) \\ &= p_{i,t} \times \mathbb{P}(r_j\theta_j(t) \leq r_i y_i, \forall j \neq i^*|\mathscr{F}_{t-1}, E_i^{\theta}(t)).\end{aligned}$$

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

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

$$\sum_{t=1}^{T}\mathbb{P}\bigg(i(t)=i, E_i^{\theta}(t), E_i^{\mu}(t)\bigg) = \sum_{j=0}^{T-1}\mathbb{E}[\frac{1}{p_{i,\tau_j+1}} - 1]. \quad (5)$$

We can upper bound the term $\mathbb{E}[\frac{1}{p_{i,\tau_j+1}}]$ in the above equation using Lemma 2 in Agrawal and Goyal's paper (see [14]) by replacing y_i in their lemma with $\frac{r_i y_i}{r_{i^*}}$: Hence, combining (5) with Lemma 2 in from [14]:

$$\begin{aligned}&\sum_{t=1}^{T}\mathbb{P}\bigg(i(t)=i, E_i^{\theta}(t), E_i^{\mu}(t)\bigg) \\ &\leq \frac{24}{\Delta_i'^2} + \sum_{j=0}^{T-1}\Theta(e^{-\Delta_i'^2\frac{j}{2}} + \frac{e^{-D_i j}}{(k+1)\Delta_i'^2} + \frac{1}{e^{\Delta_i'^2\frac{j}{4}}-1}) \\ &\leq \frac{24}{\Delta_i'^2} + \Theta(\frac{1}{\Delta_i'^2} + \frac{1}{D_i\Delta_i'^2} + \frac{1}{\Delta_i'^4}) = O(1).\end{aligned} \quad (6)$$

where $\Delta_i' = \theta_{i^*}^* - \frac{r_i y_i}{r_{i^*}}$ and $D_i = D(\frac{r_i y_i}{r_{i^*}}, \theta_{i^*}^*)$. Here, $D(a,b)$ represents the KL divergence between two Bernoulli distributions with parameters a and b respectively. We will use this notation in the rest of the paper. Therefore, we get a $O(1)$ upper bound for the first term in (3). We now consider the second term in (3), i.e., $\sum_{t=1}^{T}\mathbb{P}(i(t)=i, \overline{E_i^{\theta}(t)}, E_i^{\mu}(t))$. To analyze the second term in (3), we split the analysis into two cases:

Case 1: $\frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1$.

In this case, we have $\theta_i < x_i < y_i < \frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1$. For any $\epsilon \in (0,1]$, we choose x_i, y_i such that $D(x_i, \frac{r_{i^*}\theta_{i^*}^*}{r_i}) = \frac{D(\theta_i, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}{1+\epsilon}$ and $D(x_i, y_i) = \frac{D(x_i, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}{1+\epsilon} = \frac{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}{(1+\epsilon)^2}$. Then, using Lemma 4 in Agrawal and Goyal's paper (see [14]), we get:

$$\sum_{t=1}^{T}\mathbb{P}(i(t)=i, \overline{E_i^{\theta}(t)}, E_i^{\mu}(t)) \leq L_i(T)+1, \quad (7)$$

where $L_i(T) = \frac{\log T}{D(x_i, y_i)}$. We now consider the second case.

Case 2: $\frac{r_{i^*}\theta_{i^*}^*}{r_i} > 1$.

In this case, we have $\frac{r_i\theta_i}{r_{i^*}} < \frac{r_i x_i}{r_{i^*}} < \frac{r_i y_i}{r_{i^*}} < \theta_{i^*}^* \leq 1$. For choosing x_i, we proceed as in Case 1, i.e., for any $\epsilon \in (0,1]$ we choose x_i such that $D(\frac{r_i x_i}{r_{i^*}}, \theta_{i^*}^*) = \frac{D(\frac{r_i\theta_i}{r_{i^*}}, \theta_{i^*}^*)}{1+\epsilon}$. For selecting y_i, we pick $y_i > 1$ satisfying $\frac{r_i y_i}{r_{i^*}} < \theta_{i^*}^*$ to obtain the following lemma:

Lemma 2. *Under the conditions of Case 2, we have:*

$$\sum_{t=1}^{T}\mathbb{P}(i(t)=i, \overline{E_i^{\theta}(t)}, E_i^{\mu}(t)) = 0$$

Proof. Since under the conditions of Case 2, we choose $y_i > 1$, therefore, $\mathbb{P}(\overline{E_i^{\theta}(t)}) = \mathbb{P}(\theta_i(t) > y_i) = 0, \forall 1 \leq t \leq T$. Hence the lemma. ☐

Remark: The manner in which we handle the second term in (3) is one of the differences between the analysis here and in [14]. In Case 1, the difference between $r_i\theta_i^*$ and $r_{i^*}\theta_{i^*}^*$ is small, hence it requires more number of transmissions at r_i to distinguish it from r_{i^*}. This results in the logarithmic upper bound obtained in (7). On the other hand, in Case 2, the difference between $r_i\theta_i^*$ and $r_{i^*}\theta_{i^*}^*$ is large and hence the event $\overline{E_i^{\theta}(t)}$ happens with zero probability, resulting in Lemma 2.

Combining (7) and Lemma 2 with the fact that $L_i(T) = \frac{\log T}{D(x_i, y_i)} = (1+\epsilon)^2 \frac{\log T}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}$:

$$\begin{aligned} &\sum_{t=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^\theta(t)}, E_i^\mu(t)) \\ &\qquad \leq \mathbb{I}(\frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1)(1+\epsilon)^2 \frac{\log T}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})} \end{aligned} \tag{8}$$

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:

$$\sum_{t=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^\mu(t)}) \leq \frac{1}{D(x_i, \theta_i^*)} + 1 \tag{9}$$

Moreover, using the fact that $D(x_i, \frac{r_{i^*}\theta_{i^*}^*}{r_i}) = \frac{D(\theta_i, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}{1+\epsilon}$, after some manipulations, we can get:

$$x_i - \theta_i^* \geq \frac{\epsilon}{1+\epsilon} \times \frac{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}{\log(\frac{r_{i^*}\theta_{i^*}^*(1-\theta_i^*)}{\theta_i^*(r_i - r_{i^*}\theta_{i^*}^*)})}$$

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):

$$\sum_{t=1}^{T} \mathbb{P}(i(t) = i, \overline{E_i^\mu(t)}) \leq O(\frac{1}{\epsilon^2}) \tag{10}$$

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

$$\begin{aligned} &\mathbb{E}[N_i(T+1)] \\ &\quad \leq O(1) + \mathbb{I}(\frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1)(1+\epsilon)^2 \frac{\log T}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})} + O(\frac{1}{\epsilon^2}) \\ &\quad \leq (1+\epsilon') \frac{\mathbb{I}(\frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1) \log T}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})} + O(\frac{1}{\epsilon'^2}). \end{aligned} \tag{11}$$

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:*

$$\mathbb{E}[l(T)] \leq (1+\epsilon) \sum_{i \neq i^*} \frac{\mathbb{I}(\frac{r_{i^*}\theta_{i^*}^*}{r_i} \leq 1) \log T}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})} \Delta_i + O(\frac{n}{\epsilon^2})$$

for any $\epsilon \in (0, 1]$, where $\Delta_i = r_{i^*}\theta_{i^*}^* - r_i\theta_i^*$.

A. Discussion

The idea of decoupling the components of θ 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 θ 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 $r_i, i = 1, 2, ..., n$, 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 $Y(t) = \frac{r(t)}{r_n} X(t)$. Draw temp $\sim$ 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:*

$$\mathbb{E}[l(T)] \leq (1+\epsilon) \sum_{i \neq i^*} \frac{\log T}{D(\frac{r_i}{r_n}\theta_i^*, \frac{r_{i^*}}{r_n}\theta_{i^*}^*)} \Delta_i + O(\frac{n}{\epsilon^2})$$

for any $\epsilon \in (0, 1]$, where $\Delta_i = r_{i^*}\theta_{i^*}^* - r_i\theta_i^*$.

One of the contributions of this paper is to show that the decoupling 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 $\frac{1}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})}$, whereas Algorithm 3 has a constant factor of $\frac{1}{D(\frac{r_i}{r_n}\theta_i^*, \frac{r_{i^*}}{r_n}\theta_{i^*}^*)}$. $D(\frac{r_i}{r_n}\theta_i^*, \frac{r_{i^*}}{r_n}\theta_{i^*}^*)$ will be less than $D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})$ since the multiplication by $\frac{r_i}{r_n}$ 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 $\theta_i^* = \sum_{j=i}^{3} \nu_j^*$. 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., θ' is such that the unique optimal transmission rate for θ' is r_3 and the first two components of θ' and θ^* are the same.

Definition 6. ***(Parameters*** $X_i(s), \xi_t(i, \mathscr{F}_{t-1})$*) Let* $X_i(s)$ *denote the outcome when the rate* r_i *is transmitted for the* s^{th} *time. Let* $\xi_t(i, \mathscr{F}_{t-1})$ *denote the probability of transmitting at rate* r_i *at time* t *depending on the history until time* $t-1$*.*

To prove a lower bound on the number of times we transmit at the sub-optimal rate r_3, we need to show that probability (under θ^*) of $N_3(T+1)$ being less than a certain time-dependent threshold approaches 0 as time goes to ∞. We define this threshold to be f_T, i.e., we need to show $\mathbb{P}_{\theta^*}(N_3(T+1) \leq f_T) = 0$ as $T \to \infty$. We will choose an appropriate value of f_T later. To obtain the lower bound, we will consider a fixed but any general policy, so there will be no restriction on ξ_t. Also note that the decision making or the policy doesn't depend on θ or θ', it only depends on the history of transmission rates and their outcomes, i.e., the filtration $\mathscr{F}_{t-1}$. For ease of exposition, whenever we talk of probabilities or expectations, we will use θ^* or θ' in the subscript to clarify the probability distribution being used.

At any time t, $i(t)$ denotes the rate of transmission chosen. Therefore, until time T, we have:

$$\mathbb{P}_{\theta'}(N_3(T+1) = n) = \sum_{\mathscr{F}_T : N_3(T+1)=n} \frac{\mathbb{P}_{\theta'}(\mathscr{F}_T)}{\mathbb{P}_{\theta}(\mathscr{F}_T)} \times \mathbb{P}_{\theta}(\mathscr{F}_T)$$

Note that $\mathbb{P}_\theta(E)$ refers to the probability of an event E taking place (until the algorithm has run till time T) if the rate admissibility probability vector is θ, i.e., probability of all such filtrations $\mathscr{F}_T$ such that the event E takes place. $\mathbb{P}_\theta(E)$ will depend on the policy $\xi_t(i, \mathscr{F}_{t-1})$ but the policy itself doesn't depend on anything except the filtration $\mathscr{F}_{t-1}$. For a particular filtration $\mathscr{F}_T$ (satisfying $N_3(T+1) = n$), since the probability vectors θ^* and θ' only differ in the third component, $\mathbb{P}_{\theta^*}(\mathscr{F}_T)$ and $\mathbb{P}_{\theta'}(\mathscr{F}_T)$ will only differ due to the time slots where the rate of transmission was r_3. Let $m_3(\mathscr{F}_T)$ be the number of times transmission at r_3 resulted in a successful transmission (under $\mathscr{F}_T$). Therefore:

$$\begin{aligned}\frac{\mathbb{P}_{\theta'}(\mathscr{F}_T)}{\mathbb{P}_{\theta^*}(\mathscr{F}_T)} &= \left\{\frac{\theta'_3}{\theta^*_3}\right\}^{m_3(\mathscr{F}_T)} \times \left\{\frac{1-\theta'_3}{1-\theta^*_3}\right\}^{n-m_3(\mathscr{F}_T)} \\ &= e^{-\left(m_3(\mathscr{F}_T)\log\frac{\theta^*_3}{\theta'_3} + (n-m_3(\mathscr{F}_T))\log\frac{1-\theta^*_3}{1-\theta'_3}\right)}\end{aligned}$$

Let $L(\mathscr{F}_T) = m_3(\mathscr{F}_T)\log\frac{\theta^*_3}{\theta'_3} + (n - m_3(\mathscr{F}_T))\log\frac{1-\theta^*_3}{1-\theta'_3}$. Therefore, we have:

$$\mathbb{P}_{\theta'}(N_3(T+1) = n) = \sum_{\mathscr{F}_T : N_3(T+1)=n} e^{-L(\mathscr{F}_T)}\mathbb{P}_{\theta^*}(\mathscr{F}_T) \tag{12}$$

From (12), for the filtrations $\mathscr{F}_T$ which are likely to have similar probabilities under both θ^* and θ', the term $L(\mathscr{F}_T)$ would be small. We split the probability term in (12) into two terms by considering $L(\mathscr{F}_T) \leq c_T$ or $L(\mathscr{F}_T) > c_T$. We will choose an appropriate value of c_T later. Considering $L(\mathscr{F}_T) \leq c_T$ first:

$$\begin{aligned}&\mathbb{P}_{\theta'}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T) \\ &\quad= \sum_{\mathscr{F}_T : N_3(T+1)=n, L(\mathscr{F}_T)\leq c_T} e^{-L(\mathscr{F}_T)}\mathbb{P}_{\theta^*}(\mathscr{F}_T) \\ &\quad\geq e^{-c_T} \sum_{\mathscr{F}_T : N_3(T+1)=n, L(\mathscr{F}_T)\leq c_T} \mathbb{P}_{\theta^*}(\mathscr{F}_T) \\ &\quad= e^{-c_T}\mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T)\end{aligned}$$

We can rewrite the above inequality as:

$$\begin{aligned}&\mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T) \\ &\leq e^{c_T}\mathbb{P}_{\theta'}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T)\end{aligned} \tag{13}$$

From the law of total probability, we have:

$$\begin{aligned}&\mathbb{P}_{\theta^*}(N_3(T+1) = n) \\ &\quad= \mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T) \\ &\qquad\qquad+ \mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) > c_T) \\ &\quad\leq e^{c_T}\mathbb{P}_{\theta'}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) \leq c_T) \\ &\qquad\qquad+ \mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) > c_T) \\ &\quad\leq e^{c_T}\mathbb{P}_{\theta'}(N_3(T+1) = n) \\ &\qquad\qquad+ \mathbb{P}_{\theta^*}(\mathscr{F}_T : N_3(T+1) = n, L(\mathscr{F}_T) > c_T)\end{aligned}$$

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

$$\begin{aligned}&\mathbb{P}_{\theta^*}(N_3(T+1) \leq f_T) \\ &= \sum_{n\leq f_T} \mathbb{P}_{\theta^*}(N_3(T+1) = n) \\ &\leq e^{c_T}\sum_{n\leq f_T} \mathbb{P}_{\theta'}(N_3(T+1) = n) \\ &\qquad+ \sum_{n\leq f_T} \mathbb{P}_{\theta^*}(N_3(T+1) = n, L(\mathscr{F}_T) > c_T) \\ &= e^{c_T}\mathbb{P}_{\theta'}(N_3(T+1) \leq f_T) \\ &\qquad+ \mathbb{P}_{\theta^*}(N_3(T+1) \leq f_T, L(\mathscr{F}_T) > c_T) \\ &= e^{c_T}\mathbb{P}_{\theta'}(T - N_3(T+1) \leq T - f_T) \\ &\qquad+ \mathbb{P}_{\theta^*}(N_3(T+1) \leq f_T, L(\mathscr{F}_T) > c_T)\end{aligned} \tag{14}$$

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

$$\mathbb{P}_{\theta'}(T - N_3(T+1) \leq T - f_T) \leq \frac{\mathbb{E}_{\theta'}(T - N_3(T+1))}{T - f_T} \tag{15}$$

Under θ', 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):

$$\mathbb{P}_{\theta'}\big(T - N_3(T+1) \leq T - f_T\big) \leq o(T^{\alpha-1}) \tag{16}$$

For ease of notation, let $Y(s) = \sum_{j=1}^{s} \left\{ (X_3(j) \log(\frac{\theta_3^*}{\theta_3'}) + (1 - X_3(j)) \log(\frac{1-\theta_3^*}{1-\theta_3'}) \right\}$. Now, we consider the second term on the RHS of (14):

$$\begin{aligned}
&\mathbb{P}_{\theta^*}(N_3(T+1) \le f_T, L(\mathscr{F}_T) > c_T) \\
&= \sum_{s=1}^{f_T} \mathbb{P}_{\theta^*}(N_3(T+1) = s, L(\mathscr{F}_T) > c_T) \\
&= \sum_{s=1}^{f_T} \mathbb{P}_{\theta^*}(N_3(T+1) = s, Y(s) > c_T) \\
&\le \sum_{s=1}^{f_T} \mathbb{P}_{\theta^*}(N_3(T+1) = s, \max_{s \in 1,2,\ldots,f_T} Y(s) > c_T) \\
&= \mathbb{P}_{\theta^*}(N_3(T+1) \le f_T, \max_{s \in 1,2,\ldots,f_T} Y(s) > c_T) \\
&\le \mathbb{P}_{\theta^*}(\max_{s \in 1,2,\ldots,f_T} Y(s) > c_T)
\end{aligned}$$

where the first inequality follows from the fact that the event $\{Y(s) > c_T\} \subseteq \{\max_{s \in 1,2,\ldots,f_T} Y(s) > c_T\}, \forall 1 \le s \le f_T$. The last step follows from the fact that $\mathbb{P}(A, B) \le \mathbb{P}(A)$. Now, by Strong Law of Large Numbers, we have $\lim_{f_T \to \infty} Y(f_T) = \lim_{f_T \to \infty} \frac{1}{f_T} \sum_{s=1}^{f_T} X_{3s} \log(\frac{\theta_3}{\theta_3'}) + (1 - X_{3s}) \log(\frac{1-\theta_3}{1-\theta_3'}) = D(\theta_3^* || \theta_3')$ almost surely. Also, it is 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 $\mathbb{P}_\theta(N_{3T} \le f_T, L(\mathscr{F}_T) > c_T) \to 0$ as $f_T \to \infty$ almost surely. This takes care of the second term on the RHS in (14).

Combining (14) and (16), we observe that we need $e^{c_T} o(T^{\alpha-1}) \to 0$ as $T \to \infty$ so that $\mathbb{P}_{\theta^*}(N_3(T+1) \le f_T) = 0$ as $T \to 0$. Therefore:

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

Thus, we need $(\alpha - 1)\log T + c_T \to -\infty$ 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 $f_T = \frac{(1-\delta)c_T}{D(\theta_3^* || \theta_3')}, \delta \in (0, 1)$. These choices (of f_T, c_T) satisfy the requirements that $f_T \to \infty$ as $T \to \infty$ and that $\frac{c_T}{f_T} > D(\theta_3^* || \theta_3')$. Let $\rho(T) = \mathbb{P}_{\theta^*}(N_3(T+1) \le \frac{(1-\delta)(\frac{1-\alpha}{1+\gamma})\log T}{D(\theta_3^* || \theta_3')})$. Therefore, we conclude that:

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

(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:

$$\frac{D(\theta_3^* || \theta_3') \mathbb{E}_{\theta^*}[N_3(T+1)]}{(1-\delta)(\frac{1-\alpha}{1+\gamma}) \log T} \ge 1 - \rho(T)$$

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

$$\lim_{T \to \infty} \frac{\mathbb{E}_{\theta^*}[N_3(T+1)]}{\log T} \ge \frac{1}{D(\theta_3^* || \theta_3')} \tag{18}$$

Only thing left for us to do now is to choose an appropriate θ_3'. Note that we want $r_2\theta_2 < r_3\theta_3'$, hence we can choose any θ_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:

$$\lim_{T \to \infty} \frac{\mathbb{E}_{\theta^*}[N_3(T+1)]}{\log T} \ge \frac{1}{D(\theta_3^* || \frac{r_2}{r_3}\theta_2^*)}$$

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 $\frac{r_3\theta_3^*}{r_2} > 1$, MTS achieves $O(1)$ regret and hence we can use the trivial lower bound of 0. On the other hand, if $\frac{r_3\theta_3^*}{r_2} \le 1$, we can choose a θ_2' such that $r_2\theta_2' > r_3\theta_3^*$. The same analysis as that of Case 1 would then hold. Note that $\frac{r_2\theta_2^*}{r_3} \le 1$ is always true since $r_2 > r_3$ and $\theta_2^* \le 1$, 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:*

$$\lim_{T \to \infty} \frac{\mathbb{E}[l(T)]}{\log T} \ge \frac{\mathbb{I}(\frac{r_{i^*}\theta_{i^*}^*}{r_i} \le 1)}{D(\theta_i^*, \frac{r_{i^*}\theta_{i^*}^*}{r_i})} \Delta_i, i \ne i^*$$

where $\Delta_i = r_{i^*}\theta_{i^*}^* - r_i\theta_i^*$.

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 ν^*, rate $r_3 = 3$ is optimal. Moreover, $\frac{r_3\theta_3^*}{r_2} > 1$ and $\frac{r_3\theta_3^*}{r_1} > 1$. 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 $\nu^* = (0.3, 0, 0.7)$ (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 ν^*, rate $r_2 = 2$ is optimal. We have $\frac{r_2\theta_2^*}{r_1} > 1$, but unlike the previous experiment $\frac{r_3\theta_3^*}{r_1} \le 1$. 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.

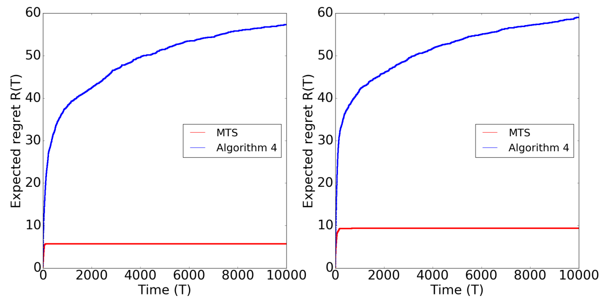

Fig. 1. Experiment 1: Implementing MTS and Algorithm 3 for $\nu^* = (0.1, 0.1, 0.8)$ (left) and $\nu^* = (0.3, 0, 0.7)$ (right). MTS achieves $O(1)$ regret for both cases while Algorithm 3 achieves logarithmic regret.

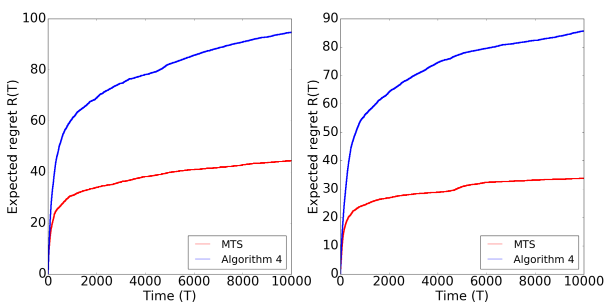

Fig. 2. Experiment 2: Implementing MTS and Algorithm 3 for $\nu^* = (0.3, 0.4, 0.3)$ (left) and $\nu^* = (0.4, 0.1, 0.5)$ (right). Both MTS and Algorithm 3 achieve logarithmic regret but MTS outperforms Algorithm 3 by a huge margin.

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

[1] "FCC Adopts Rules to Facilitate Next Generation Wireless Technologies," Federal Communications Commission, Tech. Rep., July 2016, https://www.fcc.gov/document/fcc-adopts-rules-facilitate-next-generation-wireless-technologies.

[2] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. of the IEEE*, vol. 102, pp. 366–385, 2014.

[4] Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), "Metis channel models," https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.4_v1.0.pdf, February, 2015.

[5] Qualcomm Technologies, Inc, "5G research on Waveform and Multiple Access Techniques," Tech. Rep., 2015, https://www.qualcomm.com/documents/5g-research-waveform-and-multiple-access-techniques.

[6] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications*. Cambridge University Press, 2013.

[7] IEEE Standards Association, "IEEE 802.11 standards," https://standards.ieee.org/about/get/802/802.11.html.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[9] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.

[10] M. K. Samimi, T. S. Rappaport, and G. R. MacCartney, "Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 357–360, Aug 2015.

[11] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[12] R. Aggarwal, P. Schniter, and C. E. Koksal, "Rate adaptation via link-layer feedback for goodput maximization over a time-varying channel," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4276–4285, August 2009.

[13] C. E. Koksal and P. Schniter, "Robust rate-adaptive wireless communication using ack/nak-feedback," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1752–1765, April 2012.

[14] S. Agrawal and N. Goyal, "Further optimal regret bounds for Thompson sampling," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 31. Scottsdale, Arizona, USA: PMLR, 29 Apr–01 May 2013, pp. 99–107.

[15] 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.

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

[17] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: Learning under competition," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[18] C. Tekin and M. Liu, "Approximately optimal adaptive learning in opportunistic spectrum access," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1548–1556.

[19] W. Dai, Y. Gai, and B. Krishnamachari, "Efficient online learning for opportunistic spectrum access," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 3086–3090.

[20] 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.

[21] 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.

[22] C. Li and M. J. Neely, "Network utility maximization over partially observable Markovian channels," *Performance Evaluation*, vol. 70, no. 7, pp. 528–548, 2013.