Optimal Rate Sampling in 802.11 Systems

Richard Combes†, Alexandre Proutiere†, Donggyu Yun‡, Jungseul Ok‡, Yung Yi‡

Abstract— **In 802.11 systems, Rate Adaptation (RA) is a fundamental mechanism allowing transmitters to adapt the coding and modulation scheme as well as the MIMO transmission mode to the radio channel conditions, and in turn, to learn and track the (mode, rate) pair providing the highest throughput. So far, the design of RA mechanisms has been mainly driven by heuristics. In contrast, in this paper, we rigorously formulate such design as an online stochastic optimisation problem. We solve this problem and present ORS (Optimal Rate Sampling), a family of (mode, rate) pair adaptation algorithms that provably learn as fast as it is possible the best pair for transmission. We study the performance of ORS algorithms in both stationary radio environments where the successful packet transmission probabilities at the various (mode, rate) pairs do not vary over time, and in non-stationary environments where these probabilities evolve. We show that under ORS algorithms, the throughput loss due to the need to explore sub-optimal (mode, rate) pairs does not depend on the number of available pairs, which is a crucial advantage as evolving 802.11 standards offer an increasingly large number of (mode, rate) pairs. We illustrate the efficiency of ORS algorithms (compared to the state-of-the-art algorithms) using simulations and traces extracted from 802.11 test-beds.**

I. INTRODUCTION

In 802.11 systems, transmitters select, for each packet transmission, a modulation and coding scheme as well as a MIMO mode (a diversity-oriented single-stream mode or a spatial multiplexing-oriented multiple-stream mode). Transmitters adapt the (mode, rate) pair to the channel conditions, with the objective to identify as fast as possible the pair maximising throughput, i.e., maximizing the product of the rate and of the successful packet transmission probability. The challenge in the design of (mode, rate) adaption scheme, or rate adaptation (RA) scheme for short, stems from the facts that these probabilities are unknown a priori, and that they may evolve over time.

Traditionally, RA mechanisms are based on rate sampling approaches (e.g. ARF [12], SampleRate [2]): the rate (or (mode, rate) pair) selection solely depends on the past observed packet transmission successes and failures at the various rates. As standards evolve, the number of available decisions (mode and rate pairs) gets very large, making the use of sampling approaches questionable. An alternative to sampling approaches consists in using channel measurements to predict the packet error rate (PER) under the various possible decisions. However, predicting PER accurately is difficult, and costly as measurement feedback incurs extra overhead, see e.g. [1], [5], [9], [17], [19]. As of now, it is difficult to predict whether measurement-based RA schemes will be widely adopted in the future or whether sampling approaches will continue to prevail.

In this paper, we investigate the fundamental performance limits of sampling approaches, and rigorously design the *best* sampling-based RA algorithms, i.e., maximising the expected number of packets successfully sent over a finite time horizon. These algorithms optimally explore sub-optimal decisions, and learn as fast as it is possible the best (mode, rate) pair for transmission. This contrasts with all existing RA mechansims whose design was mainly based on heuristics. Our contributions are as follows.

(i) We formulate the optimal design of RA algorithms as an online stochastic optimisation problem, referred to as a *graphically unimodal* Multi-Armed Bandit (MAB) problem (Section II).

(ii) For stationary radio environments, where the successful packet transmission probabilities using the various (mode, rate) pairs do not evolve, we derive an upper performance bound satisfied by *any* sampling-based RA algorithm. We present G-ORS, a RA algorithm applicable to 802.11 systems with single or multiple MIMO modes, and whose performance matches the upper boudn derived previously, i.e., G-ORS is optimal. As it turns out, G-ORS performance does not depend on the size of the decision space (the number of available (mode, rate) pairs), which is quite remarkable, and suggests that sampling-based RA mechanisms perform well even when the decision space is large (Section III).

(iii) For non-stationary radio environments where the successful packet transmissions do vary over time, we propose SW-G-ORS algorithm, a version of G-ORS with sliding window, and provide guarantees on its performance (Section IV).

(iv) Finally we illustrate the efficiency of our algo-

†: KTH Royal Institute of Technology, Sweden. ‡: Department of Electrical Engineering, KAIST, South Korea.

rithms using numerical experiments using both artificially generated traces, and traces extracted from test-beds (Section V).

Related work. A large array of sampling algorithms for 802.11 a/b/g systems has been proposed in the past, see e.g. [2], [12], [14]. Other sampling algorithms have been specifically developed for 802.11n MIMO systems [16], [17]. These algorithms are based on heuristics, which contrasts with the proposed schemes, developed using stochastic optimisation methods. In parallel, there has been an increasing interest for measurement-based RA algorithm, in 802.11a/b/g and n systems [6], [10], [11], [20]. Predicting PER however remain a difficult and costly task, and it is today unclear whether measurement-based schemes will be widely adopted in the future. For a comprehensive state-of-the-art on RA mechanisms, please refer to [4].

In this paper, the design of sampling-based RA algorithms is mapped into what is called a graphically unimodal MAB problem. There is an extensive literature on MAB problems, see [3] for a survey. The originality of our MAB problem lies in its structure: the average rewards achieved under the various available decisions are related. This structure is an advantage as it may be exploited to learn the best decision faster, but also brings additional theoretical challenges. Structured MAB have received little attention so far, see e.g. [3], [22]. In this paper, we provide a complete analysis of our MAB problem: we derive performance upper bound, and provide optimal sequential decision selection schemes. We also study the problem in non-stationary environments. Such environments are rarely addresed in the MAB literature, see [8], [13], [21]. As far as we know, we provide the first analysis of non-stationary and structured MAB problems.

II. Preliminaries

A. Models

We consider a single link (a transmitter-receiver pair). At time 0, the link becomes active and the transmitter has packets to send to the receiver. For each packet, the transmitter has to select a rate (for 802.11 a/b/g/ systems), or a MIMO mode and a rate (for 802.11n MIMO systems). The set of such possible decisions is denoted by $\mathcal{D}$, and is of cardinality D. The set of MIMO modes is $\mathcal{M}$ (for 802.11 a/b/g systems, there is a single available mode) and in mode m, the rate is selected from set $\mathcal{R}_m$. We write $d = (m, k)$ when the mode m is selected along with the k-th lowest rate in $\mathcal{R}_m$. Let r_d the rate selected under decision d. After a packet is sent, the transmitter is informed on whether the transmission has been successful. Based on the observed past transmission successes and failures, the transmitter has to make a decision for the next packet transmission. We denote by Π the set of all possible sequential (mode, rate) pair selection schemes. Packets are assumed to be of equal size, and without loss of generality the duration of a packet transmission at rate r is $1/r$.

1) Channel models: For the i-th packet transmission using (mode, rate) pair d, a binary random variable $X_d(i)$ represents the success ($X_d(i) = 1$) or failure ($X_d(i) = 0$) of the transmission.

Stationary radio environments. In such environments, the success transmission probabilities using the different (mode, rate) pairs do not evolve over time. This arises when the system considered is static (in particular, the transmitter and receiver do not move). Formally, $X_d(i)$, $i = 1, 2, \ldots$, are independent and identically distributed, and we denote by θ_d the success transmission probability under decision d, $\theta_d = \mathbb{E}[X_d(i)]$. Let $\mu_d = r_d \theta_d$. We denote by $d^\star$ the optimal (mode, rate) pair, $d^\star \in \arg\max_{d \in \mathcal{D}} \mu_d$.

Non-stationary radio environments. In practice, channel conditions may be non-stationary, i.e., the success probabilities could evolve over time. In many situations, the evolution over time is rather slow, see e.g. [18]. These slow variations allow us to devise sequential decision selection schemes that efficiently track the best (mode, rate) pair for transmission. In the case of non-stationary environment, we denote by $\theta_d(t)$ the success transmission probability under decision d at time t, and by $d^\star(t)$ the optimal (mode, rate) pair at time t.

Unless otherwise specified, we consider stationary radio environments. Non-stationary environments are treated in Section IV.

2) Structural properties: Our problem is to identify as fast as possible the (mode, rate) pair maximising throughput. To this aim, we leverage two crucial structural properties of the problem: (i) The successes and failures of transmissions at various (mode, rate) pairs are correlated, and (ii) in practice, we observe that the throughput vs. (mode, rate) pair function has some structure, referred to as *graphical unimodality*.

Correlations. If a transmission is successful at a high rate, it has to be successful at a lower rate, and similarly, if a low-rate transmission fails, then a transmitting at a higher rate would also fail. Formally this means that for any $m \in \mathcal{M}$, $\theta_{(m,k)} > \theta_{(m,l)}$ if $k < l$, or equivalently that $\theta = (\theta_d, d \in \mathcal{D}) \in \mathcal{T}$, where $\mathcal{T} = \{\eta \in [0,1]^D : \eta_{(m,k)} > \eta_{(m,l)}, \forall m \in \mathcal{M}, \forall k < l\}$.

Graphical unimodality. Graphical unimodality is defined through an undirected graph $G = (\mathcal{D}, E)$, whose vertices correspond to the available decisions ((mode, rate) pairs). When $(d, d') \in E$, we say that the two decisions d and d' are neighbours, and we let $\mathcal{N}(d) = \{d' \in \mathcal{D} : (d, d') \in E\}$ be the set of neighbours of d. Graphical unimodality expresses the fact that when the optimal decision is $d^\star$, then for any $d \in \mathcal{D}$, there exists a path in G from d to $d^\star$ along which the expected throughput is increased. In other words there is no *local* maximum in terms of expected throughput except at $d^\star$. Formally, $\theta \in \mathcal{U}_G$, where $\mathcal{U}_G$ is the set of parameters $\theta \in [0,1]^D$ such that, if $d^\star = \arg\max_d \mu_d$, for any $d \in \mathcal{D}$, there exists a path $(d_0 = d, d_1, \dots, d_p = d^\star)$ in G such that for any $i = 1, \dots, p$, $\mu_{d_i} > \mu_{d_{i-1}}$.

In case of 802.11 systems with a single mode, the throughput is an unimodal function of the rates, which is well known, see e.g. [17], and hence graphical unimodality holds. The corresponding graph G is a line as illustrated in Fig. 1. In 802.11n MIMO systems, we can find a graph G such that the throughput obtained at various (mode, rate) pairs is graphically unimodal with respect to G. Such a graph is presented in Fig. 1, for systems using two MIMO modes, a single-stream (SS) mode, and a double-stream (DS) mode. It has been constructed exploiting various observations and empirical results from [6], [17]. First, for a given mode (SS or DS), the throughput is unimodal in the rate. Then, when the SNR is relatively low, it has been observed that using SS mode is always better than using DS mode; this explains why for example, the (mode, rate) pair (SS,13.5) has no neighbour in the DS mode. Similarly, when the SNR is very high, then it is always optimal to use DS mode. Finally when the SNR is neither low nor high, there is no clear better mode, which explains why we need links between the two modes in the graph.

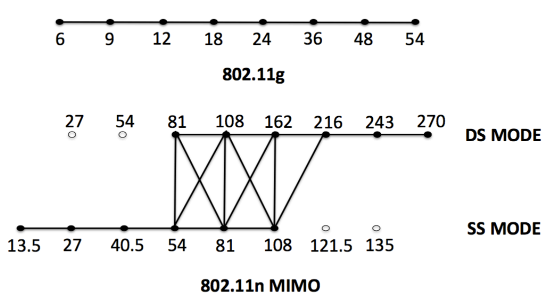


Fig. 1. Graphs G providing unimodality in 802.11g systems (above) and MIMO 802.11n systems (below). Rates are in Mbit/s. In 802.11n, two MIMO modes are considered, single-stream (SS) and double-stream (DS) modes.

B. Objectives

We now formulate the design of the best (mode, rate) pair selection algorithm as an online stochastic optimisation problem. An optimal algorithm maximises the expected number packets successfully sent over a given time horizon T. The choice of T is not really important as long as during time interval T, a large number of packets can be sent – so that inferring the success transmission probabilities efficiently is possible.

Under a given RA algorithm $\pi \in \Pi$, the number of packets $\gamma^\pi(T)$ successfully sent up to time T is: $\gamma^\pi(T) = \sum_d \sum_{i=1}^{s_d^\pi(T)} X_d(i)$, where $s_d^\pi(T)$ is the number of transmission attempts at (rate, mode) d before time T. The $s_d(T)$'s are random variables (since the rates selected under π depend on the past random successes and failures), and satisfy the following constraint: $\sum_d s_d^\pi(T) \times \frac{1}{r_d} \le T$. Wald's lemma implies that $\mathbb{E}[\gamma^\pi(T)] = \sum_d \mathbb{E}[s_d^\pi(T)]\theta_d$. Thus, our objective is to design an online algorithm solving the following stochastic optimisation problem:

$$\max_{\pi\in\Pi} \sum_d \mathbb{E}[s_d^\pi(T)]\theta_d, \tag{1}$$
$$\text{s.t. } s_d^\pi(T) \in \mathbb{N}, \forall d \in \mathcal{D} \text{ and } \sum_d s_d^\pi(T) \times \frac{1}{r_d} \le T.$$

C. Graphically Unimodal Multi-Armed Bandit

We now show that problem (1) is asymptotically (for large T) equivalent to a *graphically unimodal* MAB problem. Consider an alternative system where the duration of a packet transmission at any rate is one slot, and where decisions are taken at the beginning of each slot. When rate r is selected, and the transmission is successful, the reward is incremented by an amount of r bits. In this alternative system, the objective is to design $\pi \in \Pi$ solving the following optimisation problem.

$$\max_{\pi\in\Pi} \sum_d \mathbb{E}[t_d^\pi(T)] r_d \theta_d, \tag{2}$$
$$\text{s.t. } t_d^\pi(T) \in \mathbb{N}, \forall d \in \mathcal{D}, \text{ and } \sum_d t_d^\pi(T) \le T,$$

where $t_d^\pi(T)$ denotes the number of times decision d has been taken up to slot T. Note that if the same algorithm π is applied both in the original and alternative systems, we simply have: $t_d^\pi(T) = s_d^\pi(T)/r_d$, assuming without loss of generality that $1/r_d$ is an integer number of slots. The optimisation problem (2) corresponds to a MAB problem (see below for a formal definition). To assess the performance of $\pi \in \Pi$, it is usual in MAB literature to use the notion of *regret*. The regret up to slot T compares the performance of π to that achieved by an Oracle algorithm always selecting the best (mode, rate) pair. The regrets $R_1^\pi(T)$ and $R^\pi(T)$ of algorithm π up to time slot T in the original and alternative systems are then:

$$R_1^\pi(T) = \theta_{d^\star} \lfloor r_{d^\star} T \rfloor - \sum_d \theta_d \mathbb{E}[s_d^\pi(T)],$$
$$R^\pi(T) = \theta_{d^\star} r_{d^\star} T - \sum_d \theta_d r_d \mathbb{E}[t_d^\pi(T)].$$

In the next section, we show that for any $\pi \in \Pi$, an asymptotic lower bound of the regret $R^\pi(T)$ is of the form $c(\theta)\log(T)$ where $c(\theta)$ is a strictly positive and explicit constant. It will be also shown that there exists an algorithm $\pi \in \Pi$ that actually achieves this lower bound in the alternative system, in the sense that $\limsup_{T\to\infty} R^\pi(T)/\log(T) \le c(\theta)$. In such a case, we say that π is asymptotically optimal. The following lemma states that actually, the same lower bound holds in the original system, and that any asymptotically optimal algorithm in the alternative system is also asymptotically optimal in the original system. All proofs are presented in Appendix.

Lemma 2.1: Let $\pi \in \Pi$. For any $c > 0$, we have:

$$\left(\liminf_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \ge c\right) \Longrightarrow \left(\liminf_{T\to\infty} \frac{R_1^\pi(T)}{\log(T)} \ge c\right),$$
$$\left(\limsup_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \le c\right) \Longrightarrow \left(\limsup_{T\to\infty} \frac{R_1^\pi(T)}{\log(T)} \le c\right).$$

In view of the above lemma, instead of trying to solve (1), we can rather focus on analysing the MAB problem (2). We know that optimal algorithms for (2) will also be optimal for the original problem. Our MAB problem, whose specificity lies in its structure, i.e., in the correlations and graphical unimodality of the throughputs obtained using different (mode, rate) pairs, is summarised below.

(P_G) *Graphically Unimodal MAB.* We have a set $\mathcal{D}$ of possible decisions. If decision d is taken for the i-th time, we received a reward $r_d X_d(i)$. $(X_d(i), i = 1, 2, ...)$ are i.i.d. with Bernoulli distribution with mean θ_d. The structure of rewards across decisions are expressed through $\theta \in \mathcal{T} \cap \mathcal{U}_G$ for some graph G. The objective is to design an algorithm π minimising the regret $R^\pi(T)$ over all possible algorithms $\pi \in \Pi$.

III. Stationary Radio Environments

We consider here stationary radio environments, and first derive a lower bound on regret satisfied by *any* (mode, rate) selection algorithm. Then, we propose G-ORS (Graphical-Optimal Rate Sampling), an algorithm whose asymptotic regret matches the derived lower bound.

A. Regret lower bound

To derive a lower bound on regret for MAB problem (P_G), we first introduce the notion of *uniformly good* algorithms [15]. An algorithm π is uniformly good, if for all parameters θ, for any $\alpha > 0$, we have[1]: $\mathbb{E}[t_d^\pi(T)] = o(T^\alpha), \forall d \neq d^\star$, where $t_d^\pi(T)$ is the number of times decision d has been chosen up to time slot T, and $d^\star$ denotes the optimal decision ($d^\star$ depends on θ). Uniformly good algorithms exist as we shall see later on. We further define the following sets: for any $d \in \mathcal{D}$, $N(d) = \{d' \in \mathcal{N}(d) : \mu_d \le r_{d'}\}$. Finally, recall that the Kullback-Leibler (KL) divergence between two Bernoulli distributions with respective means p and q is: $I(p,q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$.

Theorem 3.1: Let $\pi \in \Pi$ be a uniformly good sequential decision algorithm for the MAB problem (P_G). We have:

$$\limsup_{T\to\infty} \frac{R^\pi(T)}{\log(T)} \ge c_G(\theta),$$

where

$$c_G(\theta) = \sum_{d \in N(d^\star)} \frac{r_{d^\star}\theta_{d^\star} - r_d\theta_d}{I(\theta_d, \frac{r_{d^\star}\theta_{d^\star}}{r_d})}.$$

The number of terms in the sum $c_G(\theta)$ is at most equal to the degree of the graph G. In particular, in case of 802.11 systems with a single mode, G is a line, and $c_G(\theta)$ has at most two terms. In MIMO 802.11n systems, $c_G(\theta)$ has at most 4 terms if G is the graph presented in Fig. 1. More generally, the regret lower bound does not depend on the number of available decisions, which is an important property as this number can be quite large. Note that to obtain this lower bound, the graphical unimodality of the throughput plays an important role. Indeed, without structure, i.e., assuming that $\theta \in \mathcal{T}$ only, the lower bound on regret would scale linearly with the number of available decisions, see [4] for a more detailed discussion.

B. Optimal Rate Sampling algorithm

Next we propose G-ORS, a (rate, mode) selection algorithm whose regret matches the lower bound derived in Theorem 3.1, i.e., under G-ORS, the way suboptimal (rate, mode) pairs are explored to identify the optimal pair $d^\star$ is optimal.

We denote by $t_d(n)$ the number of times decision d has been selected under G-ORS up to slot n. $\hat{\mu}_d(n) = \frac{1}{t_d(n)}\sum_{s=1}^{t_d(n)} r_d X_d(i)$ is the empirical average reward using decision d up to slot n. By convention, $\hat{\mu}_d(n) = 0$ if $t_d(n) = 0$. The leader $L(n)$ at slot n is the decision with maximum empirical average reward (ties are broken

[1] $f(T) = o(g(T))$ means that $\lim_{T\to\infty} f(T)/g(T) = 0$.

arbitrarily). Further define $l_d(n)$, the number of times decision d has been the leader up to slot n, and introduce, for any $\mu \geq 0$ and $d \in \mathcal{D}$, the sets $N(d,\mu) = \{d' \in \mathcal{N}(d) : \mu \leq r(d)\}$, and $M(d,\mu) = N(d,\mu) \cup \{d\}$. Finally, let γ be the maximum degree of a vertex in G. G-ORS algorithm assigns an index to each decision d. The index $b_d(n)$ of decision d in slot n is given by:

$$b_d(n) = \max\Big\{q \in [0, r_d] : t_d(n) I\big(\frac{\hat{\mu}_d(n)}{r_d}, \frac{q}{r_d}\big) \leq \log(l_{L(n)}(n)) + c\log(\log(l_{L(n)}(n)))\Big\}, \quad (3)$$

where c is a positive constant. For the n-th slot, G-ORS selects the decision in $M(L(n), \hat{\mu}_{L(n)}(n))$ with maximum index. Ties are broken arbitrarily.

Algorithm 1 G-ORS algorithm

For $n = 1, \ldots, D$, select (mode, rate) pair d.
For $n \geq D+1$, let $\bar{d}(n) \in \arg\max_{d \in M(L(n), \hat{\mu}_{L(n)}(n))} b_d(n)$; select (mode, rate) pair $d(n$:

$$d(n) = \begin{cases} L(n) & \text{if } (l_{L(n)}(n) - 1)/\gamma \in \mathbb{N}, \\ \bar{d}(n) & \text{otherwise .} \end{cases}$$

The next theorem states that the regret achieved under G-ORS algorithm matches the lower bound derived in Theorem 3.1.

Theorem 3.2: Fix $\theta \in \mathcal{T} \cap \mathcal{U}_G$. For all $\epsilon > 0$, under algorithm $\pi =$ G-ORS, the regret at time T is bounded by:

$$R^\pi(T) \leq (1+\epsilon) c_G(\theta) \log(T) + O(\log(\log(T))).$$

As a consequence:

$$\limsup_{T \to \infty} \frac{R^\pi(T)}{\log(T)} \leq c_G(\theta).$$

It is worth noting again that the regret of G-ORS does not depend on the size of the decision space, which constitutes a crucial property as the decision space gets larger as new standards appear.

IV. Non-stationary Radio Environments

In this section, we consider non-stationary radio environments where the transmission success probabilities $\theta(t)$ at various (mode, rate) pairs evolve over time. Based on G-ORS algorithm, we design SW-G-ORS (SW stand for Sliding Window) algorithm that efficiently tracks the best mode and rate for transmission in non-stationary environments, provided that the speed at which $\theta(t)$ evolves remains controlled. To simplify the presentation, we present the algorithm in the alternative system (see Section II), where time is slotted, and at the beginning of each slot, a (mode, rate) pair is selected, i.e., we study non-stationary versions of MAB problem (P_G).

We denote by $X_d(t)$ the binary r.v. indicating the success or failure of a transmission using (mode, rate) pair d at the t-th slot. $(X_d(t), t = 1, 2, \ldots)$ are independent with evolving mean $\theta_d(t) = \mathbb{E}[X_d(t)]$. The objective is to design a sequential decision scheme minimising the regret $R^\pi(T)$ over all possible algorithms $\pi \in \Pi$, where

$$R^\pi(T) = \sum_{t=1}^{T} \left(\mu_{d^\star(t)}(t) - \mu_{d^\pi(t)}(t)\right),$$

and $d^\star(t)$ (resp. $d^\pi(t)$) denotes the best decision (resp. the decision selected under π) at time t. $d^\star(t) = \arg\max_d \mu_{d(t)}(t)$. The above definition of regret is not standard: the regret is exactly equal to 0 only if the algorithm would be aware of the best transmission decision at any time. This notion of regret really quantifies the ability of the algorithm π to track the best decision. In particular, as shown in [7], under some mild assumptions on the way $\theta(t)$ varies, we cannot expect to obtain a regret that scales sublinearly with time horizon T. The regret is linear, and what we really wish to minimise is the regret per unit time $R(T)/T$.

A. SW-G-ORS algorithm

A natural and efficient way of tracking the changes of $\theta(t)$ over time is to select a decision at time t based on observations made over a fixed time window preceding t, i.e., to account for transmissions that occurred between time $t-\tau$ and t, see e.g. [7]. The time window τ is chosen empirically: it must be large enough (to be able to learn), but small enough so that the channel conditions do not vary significantly during a period of duration τ. In SW-G-ORS algorithm, we apply this idea. Let $d(t)$ denote the index of the (mode, rate) pair selected at time t. The empirical average reward under decision d at time n over a window of size $\tau+1$ is: $\hat{\mu}_d^\tau(n) = \frac{1}{t_d^\tau(n)} \sum_{t=n-\tau}^{n} r_d X_d(t) \mathbb{1}\{d(t) = d\}$, where $t_d^\tau(n) = \sum_{t=n-\tau}^{n} \mathbb{1}\{d(t) = d\}$.

Based on $\hat{\mu}_d^\tau(n)$, we can redefine as previously $L^\tau(n)$, the leader at time n, $l_d^\tau(n) = \sum_{t=n-\tau}^{n} \mathbb{1}\{L^\tau(t) = d\}$, the number of times d has been the leader over the window τ preceding n, and $b_d^\tau(n)$, the index of decision d at time n (see definition (3) where all quantities are considered over sliding time windows). We give the pseudo-code of SW-G-ORS below:

Algorithm 2 SW-G-ORS with window size $\tau+1$

For $n = 1, \dots, D$, select the rate with index $d(n) = n$. For $n = D+1, \dots,$ select decision $d(n)$ where: $d(n) = L^\tau(n)$ if $(l^\tau_{L^\tau(n)}(n) - 1)/\gamma \in \mathbb{N}$, and
$d(n) = \arg\max_{d \in M(L^\tau(n), \hat{\mu}^\tau_{L^\tau(n)}(n))} b^\tau_d(n)$ otherwise.

B. Regret analysis

To analyse the performance of SW-G-ORS, we make the following assumptions. $\theta(t)$ varies over time in a smooth way, i.e., for any d, $\theta_d(t)$ is σ-lipschitz: $|\theta_d(t') - \theta_d(t)| \leq \sigma|t' - t|$. We further assume that graphical unimodality holds at all time, in the sense that for any t, $\theta(t) \in \mathcal{T} \cap \bar{\mathcal{U}}_G$, where $\bar{\mathcal{U}}_G$ is the smallest closed set containing $\mathcal{U}_G$ (taking the closure of $\mathcal{U}_G$ is needed: the optimal decision changes, and hence at some times, two decisions my have the same average throughput). Finally, we assume that the proportion of time where two decisions are not well separated (they have similar throughput) is controlled in the following sense: there exists Δ_0 and $C > 0$ such that for any $\Delta \leq \Delta_0$, for any d and $d' \in N(d)$,

$$\frac{1}{T}\sum_{n=1}^{T} 1_{\{|r_k\theta_k(n) - r_{k^\star(n)}\theta_{k^\star}(n)| \leq \Delta\}} \leq C \times \Delta + o(T). \quad (4)$$

This assumption is natural, and typically hold in practice: C upper-bounds the proportion of time when throughputs under d and d' cross each other (in MAB, it is in general problematic to have decisions with very similar average rewards).

Theorem 4.1: Under the above assumptions, the regret under $\pi =$SW-G-ORS satisfies:

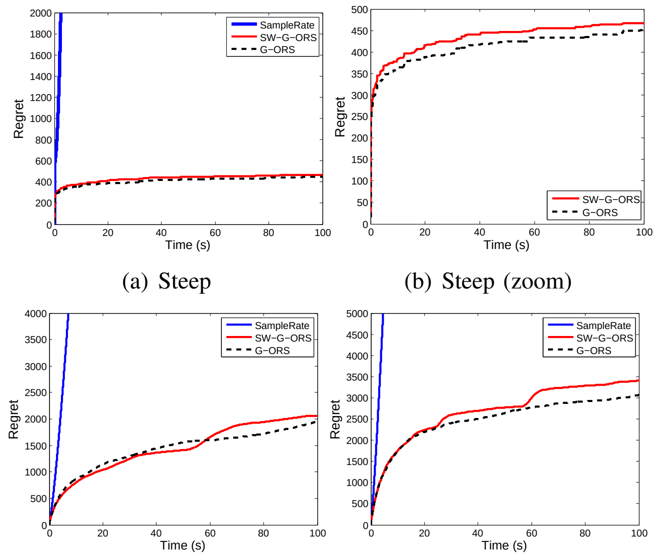
$$\limsup_{T\to\infty} \frac{R^\pi(T)}{T} \leq c'_G \sigma^{2/5} \log(1/\sigma),$$

where $c'_G(\theta)$ depends on C, and the graph G.

Note that $\sigma^{2/5}\log(1/\sigma)$ tends to 0 as $\sigma \to 0$, which indicates that the regret per unit time vanishes when we slow down the evolution of $\theta(t)$, i.e., SW-G-ORS tracks the best decision if $\theta(t)$ evolves slowly. Also observe again that the regret upper bound does not depend on the size of decision space.

V. Numerical Experiments

In this section, we illustrate the efficiency of our algorithms using traces that are either artificially generated or extracted from test-beds. Artificial traces allow us to build a performance benchmark including various kinds of radio channel scenarios as those used in [2].



(a) Steep (b) Steep (zoom)

(c) Gradual (d) Lossy

Fig. 2. Regret vs. time in stationary environments under SampleRate, G-ORS, and SW-G-ORS.

A. Artificial traces

1) 802.11g: We first consider 802.11 systems with a single MIMO mode: we use 802.11g standard with 8 available rates from 6 to 54 Mbit/s. Algorithms are tested in three different scenarios as in [2]: *steep*, *gradual*, and *lossy*. In steep scenario, the successful transmission probability is either very high or very low. In gradual scenario, the best rate is the highest rate with success probability higher than 0.5. Finally in lossy scenario, the best rate has a low success probability, i.e., less than 0.5. In stationary environments, the success transmission probabilities at the various rates are (steep) $\theta = (.99, .98, .96, .93, 0.9, .1, .06, .04)$, (gradual) $\theta = (.95, .9, .8, .65, .45, .25, .15, .1)$, and (lossy) $\theta = (.9, .8, .7, .55, .45, .35, .2, .1)$. Observe that in all cases, $\theta \in \mathcal{T} \cap \mathcal{U}_G$ (unimodality holds). We compare G-ORS and SW-G-ORS to SampleRate, where the size of sliding window is taken equal to 10s. SampleRate explores new rates every ten packet transmissions, and hence has a regret linearly increasing with time. G-ORS and SW-G-ORS explore rates in an optimal manner, and significantly outperform SampleRate – see Fig. 2.

For non-stationary environments, we artificially generate varying success probabilities $\theta(t)$ as depicted Fig. **??**. At the beginning, the value of θ corresponds to a steep scenario. It then evolves to a gradual and finally lossy scenario. Fig. **??** compares the performance of SW-G-ORS to that of SampleRate and of an oracle algorithm (that always knows the best rate for transmission). SW-G-ORS again outperforms SampleRate, and is not far from the Oracle algorithm.

2) MIMO 802.11n:

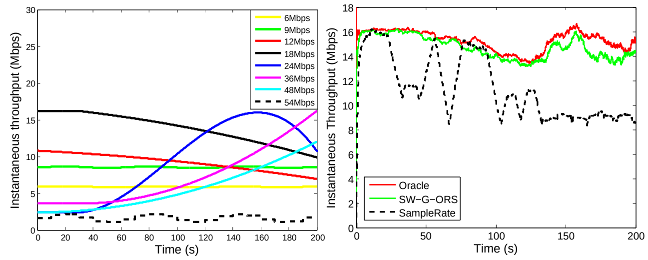
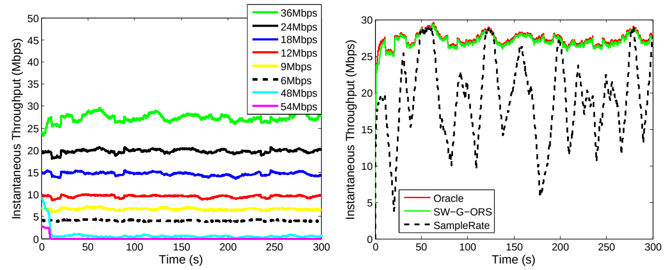


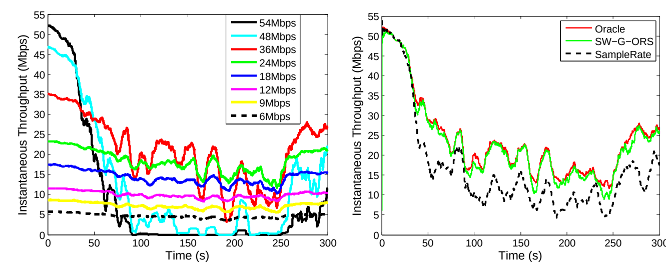
Fig. 3. Artificially generated non-stationary environment: (left) throughput at different rates; (right) throughput (averaged over .5s) under SW-G-ORS, SampleRate, and the Oracle algorithm.

B. Test-bed traces

The traces used here come from indoor 802.11g test-bed, and from the 802.11n test-bed used [6].



(a) Stationary environment

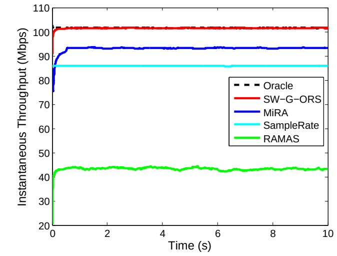


(b) Non-Stationary environment

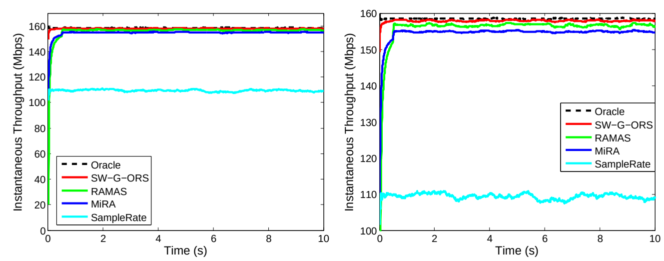
Fig. 4. 802.11g test-bed traces. Throughput evolution at different rates (left), and throughput under SW-G-ORS, SampleRate, and the Orcale algorithm (right) in stationary (top) and non-stationary (bottom) environment.

1) 802.11g:

2) MIMO 802.11n:



(a) Artificial traces



(b) Test-bed traces

Fig. 5. Stationary environment

VI. Conclusion

References

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11 b mesh network. *ACM SIGCOMM Computer Communication Review*, 34(4):121–132, 2004.

[2] J. Bicket. *Bit-rate selection in wireless networks*. PhD thesis, Massachusetts Institute of Technology, 2005.

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

[4] R. Combes, A. Proutiere, D. Yun, J. Ok, and Y. Yi. Optimal rate sampling in 802.11 systems. *arxiv.org*, 2013.

[5] R. Crepaldi, J. Lee, R. Etkin, S.-J. Lee, and R. Kravets. Csi-sf: Estimating wireless channel state using csi sampling amp; fusion. In *INFOCOM, 2012 Proceedings IEEE*, pages 154–162, 2012.

[6] L. Deek, E. Garcia-Villegas, E. Belding, S.-J. Lee, and K. Almeroth. Joint rate and channel width adaptation in 802.11 mimo wireless networks. In *Proceedings of IEEE Secon*, 2013.

[7] A. Garivier and E. Moulines. On upper-confidence bound policies for non-stationary bandit problems, 2008. ArXiv e-print. http://arxiv.org/abs/0805.3415.

[8] A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *Proceedings of the 22nd international conference on Algorithmic learning theory*, ALT'11, pages 174–188, 2011.

[9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. *SIGCOMM Comput. Commun. Rev.*, 40(4):159–170, Aug. 2010.

[10] G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive mac protocol for multi-hop wireless networks. In *Proceedings of ACM Mobicom*, 2001.

[11] G. Judd, X. Wang, and P. Steenkiste. Efficient channel-aware rate adaptation in dynamic environments. In *Proceedings of ACM MobiSys*, 2008.

[12] A. Kamerman and L. Monteban. Wavelan-ii: a high-performance wireless lan for the unlicensed band. *Bell Labs technical journal*, 2(3):118–133, 1997.

[13] L. Kocsis and C. Szepesvári. Discounted ucb. In *Proceedings of the 2dn PASCAL Challenges Workshop*, 2006.

[14] M. Lacage, M. Manshaei, and T. Turletti. Ieee 802.11 rate adaptation: a practical approach. In *Proceedings of MSWiM*, pages 126–134. ACM, 2004.

[15] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–2, 1985.

[16] D. Nguyen and J. Garcia-Luna-Aceves. A practical approach to rate adaptation for multi-antenna systems. In *Proceedings of IEEE ICNP*, 2011.

[17] I. Pefkianakis, Y. Hu, S. H. Wong, H. Yang, and S. Lu. Mimo rate adaptation in 802.11n wireless networks. In *Proceedings of ACM Mobicom*, 2010.

[18] B. Radunovic, A. Proutiere, D. Gunawardena, and P. Key. Dynamic channel, rate selection and scheduling for white spaces. In *Proceedings of ACM CoNEXT*, 2011.
[19] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. *SIGCOMM Comput. Commun. Rev.*, 36(4):51–62, Aug. 2006.
[20] B. Sagdehi, V. Kanodia, A. Sabharwal, and E. Knightly. Opportunistic media access for multirate ad hoc networks. In *Proceedings of ACM Mobicom*, 2002.
[21] A. Slivkins and E. Upfal. Adapting to a changing environment: the brownian restless bandits. In *COLT*, pages 343–354, 2008.
[22] J. Y. Yu and S. Mannor. Unimodal bandits. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 41–48, New York, NY, USA, 2011. ACM.

Appendix

Proof of Lemma 2.1

Let $T > 0$. By time T, we know that there have been at least $\lfloor Tr_1 \rfloor$ transmissions, but no more than $\lceil Tr_K \rceil$. Also observe that both regrets R^π and R_1^π are increasing functions of time. We deduce that:

$$R^\pi(\lfloor Tr_1 \rfloor) \le R_1^\pi(T) \le R^\pi(\lceil Tr_K \rceil).$$

Now

$$\begin{aligned}\liminf_{T\to\infty} \frac{R_1^\pi(T)}{\log(T)} &\ge \liminf_{T\to\infty} \frac{R^\pi(\lfloor Tr_1 \rfloor)}{\log(T)} \\ &= \liminf_{T\to\infty} \frac{R^\pi(\lfloor Tr_1 \rfloor)}{\log(\lfloor Tr_1 \rfloor)} \ge c.\end{aligned}$$

The second statement can be derived similarly.