Optimal Rate Sampling in 802.11 Systems

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

***Abstract*—Rate Adaptation (RA) is a fundamental mechanism in 802.11 systems. It allows 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 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. This 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., maximising the product of the rate and the successful packet transmission probability. The challenge in the design of (mode, rate) adaptation schemes, or rate adaptation (RA) schemes 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 [13], SampleRate [2]): the rate (or (mode, rate) pair) selection solely depends on the past observed packet transmission successes and failures. 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], [10], [18], [20]. 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 mechanisms whose design has mainly been based on heuristics. We present the following contributions.

(i) We formulate the design of optimal 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 (Graphical-Optimal Rate Sampling), a RA algorithm applicable to 802.11 systems with single or multiple MIMO modes, and whose performance matches the upper bound derived previously. Thus, G-ORS is optimal. As it turns out, the performance of G-ORS 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, a version of G-ORS with a sliding window, and provide guarantees on its performance (Section IV).

(iv) Finally we illustrate the efficiency of our algorithms through numerical experiments using both artificially generated traces, and traces extracted from test-beds (Section V).

Related work. A large array of sampling-based RA algorithms for 802.11 a/b/g systems has been proposed in the past, see e.g. [2], [13], [15]. Other sampling algorithms have been specifically developed for 802.11n MIMO systems [17], [18]. These algorithms are based on heuristics, contrary to the proposed schemes that are developed using stochastic optimisation methods. In parallel, there has been an increasing interest for measurement-based RA algorithms, in 802.11a/b/g/n systems [6], [11], [12], [21]. Predicting PER however remains a difficult and costly task, and today, very few devices allow measurement feedback. For a comprehensive survey on RA mechanisms, please refer to [4].

In this paper, the design of sampling-based RA algorithms is mapped into a so-called graphically unimodal MAB problem. The connection between MAB and RA algorithms has been mentioned in [19] (but the authors of [19] do not solve their MAB problem, and they present algorithms based on heuristics only). There is an extensive literature on MAB problems, see [3] for a survey. The originality of our problem lies in its structure: the average rewards achieved under the

†: KTH Royal Institute of Technology, Sweden. ×: INRIA / ENS, Paris France. ‡: Department of Electrical Engineering, KAIST, South Korea.
978-1-4799-3360-0/14/$31.00

various available decisions are related. This structure is an advantage as it may be exploited to learn the best decision faster, but it also brings additional theoretical challenges. Structured MAB problems have received little attention so far, see e.g. [3], [23]. In this paper, we provide a complete analysis of our MAB problem: we derive a performance upper bound, and provide optimal sequential decision schemes. We also study the problem in non-stationary environments. Such environments are rarely addressed in the MAB literature, see [8], [14], [22]. As far as we know, this paper provides 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 $\mathcal{R}_m$. For $d \in \mathcal{D}$, 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 be 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 unit size (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 $X_d(i)$ indicates 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, $X_d(i)$, $i = 1, 2, \dots$, 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. [19]. These slow variations allow us to devise RA algorithms that efficiently track the best (mode, rate) pair for transmission. In the case of non-stationary environments, we denote by $\theta_d(t)$ the success transmission probability under decision d, 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 with the highest 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 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 means 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 the 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. [18], 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], [18]. 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) pairs (DS,27) and (DS,54) are initially removed from the graph. 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 mode that clearly outperforms the other, which explains why we need edges between the two modes in the graph.

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 an interval of duration 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

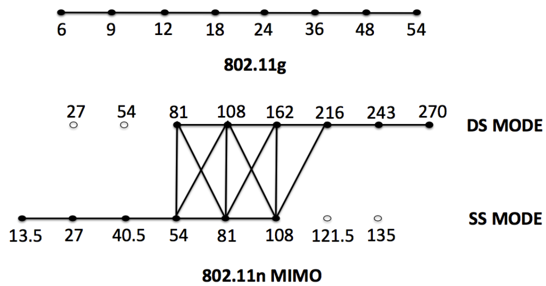


Fig. 1. Graphs G providing unimodality in 802.11g systems (above) and MIMO 802.11n systems (below). Rates are in Mbit/s. In 802.11n, two MIMO modes are considered, single-stream (SS) and double-stream (DS) modes.

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. 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 the 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 constant. It will also be 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 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 postponed to the 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 receive 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 [16]. An algorithm π is uniformly good, if for all parameters θ, for any $\alpha > 0$, we have[1]: $\mathbb{E}[t_d^\pi(T)] = o(T^\alpha), \forall d \ne d^\star$, 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, for any $d \in \mathcal{D}$, the set $N(d) = \{d' \in \mathcal{N}(d) : \mu_d \le r_{d'}\}$. 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:

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

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

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 (only assuming that $\theta \in \mathcal{T}$), the lower bound on regret scales 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 best 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, and by $\hat{\mu}_d(n) = \frac{1}{t_d(n)} \sum_{i=1}^{t_d(n)} r_d X_d(i)$ 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 arbitrarily). Further define $l_d(n)$, the number of times decision d has been the leader up to slot n, and introduce, for any $d \in \mathcal{D}$, the set $M(d) = \mathcal{N}(d) \cup \{d\}$. 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:

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

where $c \ge 3$ is a positive constant. For the n-th slot, G-ORS essentially selects the decision in the neighbourhood of the leader with maximal index. Ties are broken arbitrarily. More precisely:

Algorithm 1 G-ORS algorithm

For $n = 1, \dots, D$, select (mode, rate) pair $d(n) = n$.
For $n = D+1 \dots$, let $\bar{d} \in \arg\max_{d \in M(L(n))} b_d(n)$; select pair $d(n)$ with

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

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

Theorem 3.2: Fix $\theta \in \mathcal{T} \cap \mathcal{U}_G$. The regret of algorithm π = G-ORS satisfies:

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

It is worth noting again that the regret of G-ORS does not depend on the size of the decision space, which, as already mentioned, constitutes a highly desirable property. In the proof of the above theorem, we actually get a more precise bound on $R^\pi(T)$: it is shown that for any $\epsilon > 0$, $R^\pi(T) \le (1+\epsilon)c_G(\theta)\log(T) + O(\log(\log(T)))$, see [4] for a complete discussion.

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, we design the SW-G-ORS (SW stands for Sliding Window) algorithm that efficiently tracks the best mode and rate for transmission, 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. $(X_d(t), t = 1, 2, \dots)$ are independent with evolving mean $\theta_d(t) = \mathbb{E}[X_d(t)]$. Let $\mu_d(t) = r_d\theta_d(t)$. The objective is to design a sequential decision scheme minimising the *non-stationary* regret $R^\pi_{\text{NS}}(T)$ over all possible algorithms $\pi \in \Pi$, where

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

and $d^\star(t)$ (resp. $d^\pi(t)$) denotes the best decision (resp. the decision selected under π) at time t. $d^\star(t) = \arg\max_d \mu_d(t)$. The above definition of regret is not standard: the regret is exactly equal to 0 only if the best transmission decision is known at each time. This notion of regret really quantifies the ability of the algorithm π to track the best decision. In particular, as shown in [8], under some mild assumptions on the way $\theta(t)$ varies, we cannot expect to obtain a regret that scales sub-linearly with time horizon T. The regret is linear, and what we really wish to minimise is the regret per unit of time $R^\pi_{\text{NS}}(T)/T$.

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 slots $t - \tau$ and t, see e.g. [8]. The time window τ is chosen empirically: it must be large enough (to accurately estimate throughputs), but small enough so that the channel conditions do not vary significantly during a period of duration τ. We apply this idea to design SW-G-ORS. Let $d(t)$ denote the (mode, rate) pair selected at time t. The empirical average throughput under decision d at time n over a window of size $\tau + 1$ is: $\hat{\mu}^\tau_d(n) = \frac{1}{t^\tau_d(n)} \sum_{t=n-\tau}^{n} r_d X_d(t) \ \{d(t) = d\}$, where $t^\tau_d(n) = \sum_{t=n-\tau}^{n} \ \{d(t) = d\}$. We use the convention that $\hat{\mu}^\tau_d(n) = 0$ if $t^\tau_d(n) = 0$. Based on $\hat{\mu}^\tau_d(n)$, we can redefine as previously $L^\tau(n)$, the leader at time n, $l^\tau_d(n) = \sum_{t=n-\tau}^{n} \ \{L^\tau(t) = d\}$, the number of times d has been the leader over the window τ preceding n, and $b^\tau_d(n)$, the index of decision d at time n (in definition (3), all quantities are considered over sliding time windows). The pseudo-code of SW-G-ORS is given below.

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

For $n = 1, \ldots, D$, select (mode, rate) pair $d(n) = n$.
For $n = D + 1 \ldots$, let $\bar{d} \in \arg\max_{d \in M(L^\tau(n))} b_d^\tau(n)$; select pair $d(n)$ with

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

B. Regret analysis

To analyse the performance of SW-G-ORS, we make the following assumptions. $\theta(t)$ varies smoothly over time. For any d, $\theta_d(t)$ is σ-Lipschitz (i.e., $|\theta_d(t') - \theta_d(t)| \le \sigma |t' - t|$). 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 may 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 $\Delta \le \Delta_0$, for any d and $d' \in N(d)$,

$$\lim\sup_{T \to \infty} \frac{1}{T} \sum_{n=1}^{T} 1_{\{|\mu_d(n) - \mu_{d'}(n)| \le \Delta\}} \le C\Delta.$$

This assumption is natural, and typically holds in practice: the constant C characterises the proportion of time when the throughputs achieved under d and d' cross each other. In MAB problems, it is in general problematic to have decisions with very similar average rewards, and this constitutes the main difficulty in the regret analysis in non-stationary environments where rewards of various decisions cross each other.

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

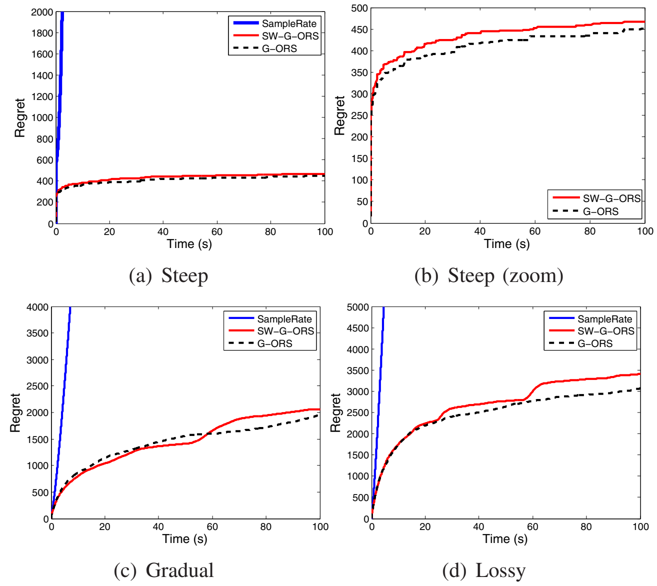
$$\lim\sup_{T \to \infty} \frac{R^\pi_{\text{NS}}(T)}{T} \le C' \sigma^{1/4} \log(1/\sigma),$$

where the constant $C' > 0$ is uniform in σ.

Note that $\sigma^{1/4} \log(1/\sigma)$ 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 that the performance guarantee on SW-G-ORS does not depend on the size of the decision space (i.e., on D). The proof of the above theorem is long and technical, and is omitted here. It is presented in details in [4].

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]. They also provide the opportunity to create non-stationary radio environments (in the literature, RA mechanisms are mostly evaluated in stationary environments).



(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.Unit of regret: 0.5kbits

A. 802.11g systems

1) Artificial traces: We first consider 802.11g 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 scenarios, the successful transmission probability is either very high or very low. In gradual scenarios, the best rate is the highest rate with success probability higher than 0.5. Finally in lossy scenarios, 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 the 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 explores rates in an optimal manner, and significantly outperforms SampleRate as demonstrated see Fig. 2.

For non-stationary environments, we artificially generate varying success probabilities $\theta(t)$ as depicted in Fig. 3 (left). At the beginning, the value of θ corresponds to a steep scenario. It then evolves to a gradual and finally lossy scenario. Fig. 3 (right) 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 its performance is close to that of the Oracle algorithm.

2) Test-bed traces: We now present results obtained in our 802.11g test-bed, consisting of two 802.11g nodes (SparkLAN WPEA 123AG/E) connected in ad-hoc mode. We collect traces recording the throughputs at the 8 available rates, and then use these traces to test SampleRate, and SW-G-ORS algorithm. Packets are of size 1500 bytes. We generate two kinds of traces: (a) when the two nodes have fixed positions, the successful packet transmission probabilities are roughly constant – we have a stationary environment; (b) the receiver

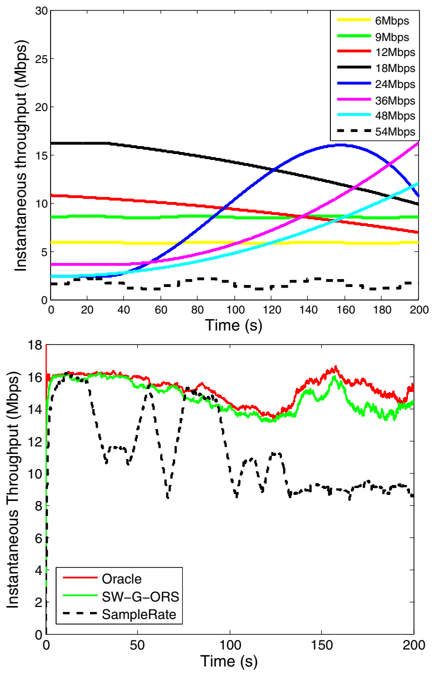
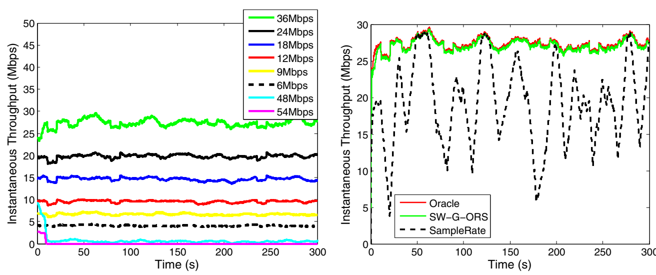
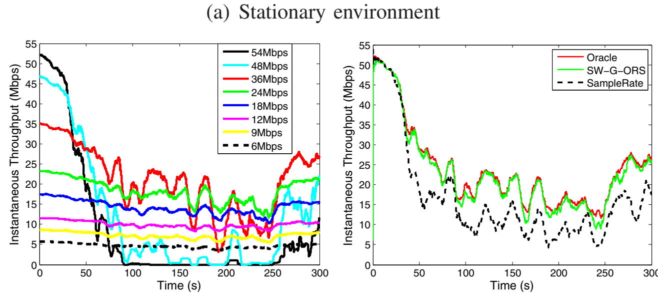


Fig. 3. Artificially generated non-stationary environment: (above) throughput at different rates; (below) throughput (averaged over 10s) under SW-G-ORS, SampleRate, and the Oracle algorithm.

is then moved (at the speed of a pedestrian), generating a non-stationary environment. The results are presented in Fig. 4. Again in both scenarios, SW-G-ORS clearly outperforms SampleRate, and exhibits a performance close that of the Oracle algorithm.



(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 Oracle algorithm (right) in stationary (top) and non-stationary (bottom) environment.

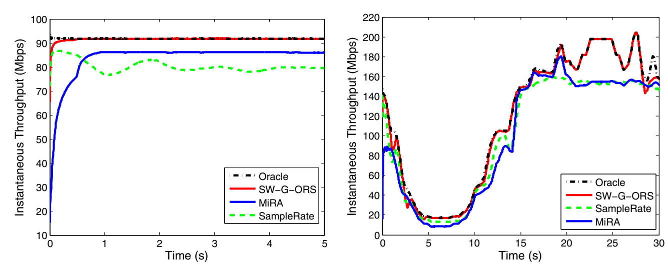
B. 802.11n MIMO systems

Next we investigate the performance of SW-G-ORS in 802.11n MIMO systems with two modes, SS and DS, as in [6], [18]. We use frame aggregation (30 packets per frame) which is essential in 802.11n. SW-G-ORS is compared to MiRA [18] and SampleRate. To define SW-G-ORS, we use the graph G depicted in Fig. 1. The sliding window for SW-G-ORS and SampleRate is taken equal to 1s. MiRA is a RA algorithm specifically designed for MIMO systems. It zigzags between MIMO modes to find the best (mode, rate) pair. In its implementation, we use, as suggested in [18], the following parameters: $\alpha = 1/8$, $\beta = 1/4$, and $T_0 = 0.2ms$.

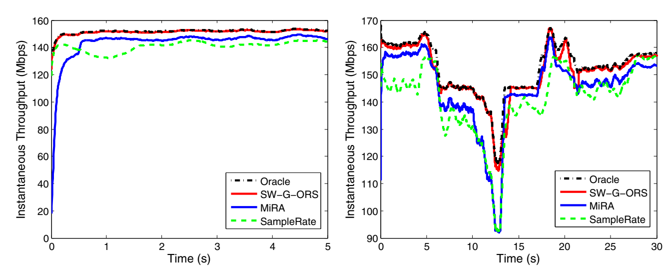
Artificial traces. To generate artificial traces, we use results from [6], and more specifically, the mapping between channel measurements (the SNR and diffSNR[2], see [6]) and the packet transmission success probabilities. For non-stationary scenarios, we artificially vary (smoothly) the SNR and diffSNR, and then deduce the corresponding evolution of the PER at various (mode, rate) pairs.

Test-bed traces. We also exploit real test-bed traces extracted from [6]. These traces correspond to stationary environments, and to generate non-stationary traces, we let the system evolve between 5 stationary scenarios.

Results are presented in Fig. 5. Instantaneous throughputs are computed on a window of size 0.5s. In stationary environments, we observe, as expected, that SW-G-ORS is able to learn the best (mode, rate) pair very rapidly, faster than any other algorithm. In the tested scenarios, we find that both MiRA and SampleRate were also able to find the best pair (there are scenarios where SampleRate is not able to do it [18]). Note however that SW-G-ORS provides a better throughput than MiRA and SampleRate (these algorithms do not explore (mode, rate) pairs in an optimal way). In non-stationary scenarios, the throughput of SW-G-ORS is really close to that of the Oracle algorithm. MiRA and SampleRate do seem to be able to track the best (mode, rate) pair, but the performance loss compared to SW-G-ORS can be quite significant.



(a) Artificial traces



(b) Test-bed traces

Fig. 5. Throughput under SW-G-ORS, SampleRate, MiRA, and the Oracle algorithm in 802.11n systems: (left) stationary environments, (right) non-stationary environments; (top) artificial traces, (bottom) test-bed traces.

[2] The maximal gap between the SNRs measured at the various antennas.

VI. CONCLUSION

In this paper, we investigated the fundamental limits of sampling approaches for the design of RA adaptation algorithms in 802.11 systems. We developed G-ORS, an algorithm that provably learns as fast as it is possible the best MIMO mode and rate for transmission. The proposed design methodology is based on online stochastic optimisation techniques: it is versatile, and can be easily adapted to evolving 802.11 standards. We also proposed SW-G-ORS, an extension of G-ORS for non-stationary radio environments. Our numerical experiments showed that G-ORS and SW-G-ORS outperform state-of-the-art sampling-based RA algorithms. This is not surprising as our algorithms are by design optimal. This performance superiority is due to the fact that under G-ORS, the way sub-optimal mode and rate pairs are explored is carefully and optimally controlled.

ACKNOWLEDGEMENTS

We would like to thank the authors of [6] for sharing their test-bed traces. R. Combes' and A. Proutiere's work is supported by the ERC 308267 FSA grant, Vetenskapradet and SSF. This work was also supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government(MSIP)(2013R1A2A2A01067633).

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 and fusion. In *Proceedings of IEEE INFOCOM*, 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 O. Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of Conference On Learning Theory (COLT)*, 2011.

[8] A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems. In *Proceedings of ALT*, pages 174–188, 2011.

[9] T. L. Graves and T. L. Lai. Asymptotically efficient adaptive choice of control laws in controlled markov chains. *SIAM J. Control and Optimization*, 35(3):715–743, 1997.

[10] 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.

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

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

[13] 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.

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

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

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

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

[18] 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.

[19] B. Radunovic, A. Proutiere, D. Gunawardena, and P. Key. Dynamic channel, rate selection and scheduling for white spaces. In *Proceedings of ACM CoNEXT*, 2011.

[20] 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.

[21] B. Sagdehi, V. Kanodia, A. Sabharwal, and E. Knightly. Opportunistic media access for multirate ad hoc networks. In *Proceedings of ACM Mobicom*, 2002.

[22] A. Slivkins and E. Upfal. Adapting to a changing environment: the brownian restless bandits. In *COLT*, pages 343–354, 2008.

[23] J. Y. Yu and S. Mannor. Unimodal bandits. In *Proceedings of ICML*, 2011.

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.

PROOF OF THEOREM 3.1

We apply the theory of controlled Markov chains developed in [9]. Using the same terminology and notation as in [9], the parameter θ takes values in $\mathcal{T} \cap \mathcal{U}_G$; the Markov chain has values in $\mathcal{S} = \{0, r_1, \dots, r_D\}$; the set of control laws is $\{1, \dots, D\}$, i.e., each control law corresponds to a (mode, rate) pair; the transition probabilities are given as follows: for all $x, y \in \mathcal{S}$,

$$p(x, y; d, \theta) = p(y; d, \theta) = \begin{cases} \theta_d, & \text{if } y = r_d, \\ 1 - \theta_d, & \text{if } y = 0; \end{cases}$$

finally, the reward $r(x, d) = x$.

We now fix $\theta \in \mathcal{T} \cap \mathcal{U}_G$. Define $I^d(\theta, \lambda) = I(\theta_d, \lambda_d)$ for any d. Further define the set $B(\theta)$ consisting of all *bad* parameters $\lambda \in \mathcal{T} \cap \mathcal{U}_G$ such that $d^\star$ is not optimal under parameter λ, but which are statistically *indistinguishable* from θ:

$$B(\theta) = \{\lambda \in \mathcal{T} \cap \mathcal{U}_G : \lambda_{d^\star} = \theta_{d^\star}, \max_d r_d \lambda_d > r_{d^\star} \lambda_{d^\star}\},$$

$B(\theta)$ can be written as the union of sets $B_d(\theta)$, $d = 1, \dots, D$ defined as: $B_d(\theta) = \{\lambda \in B(\theta) : r_d\lambda_d > r_{d^\star}\lambda_{d^\star}\}$. Note that $B_d(\theta) = \emptyset$ if $r_d < r_{d^\star}\theta_{d^\star}$. Define $P = \{d : r_d \geq r_{d^\star}\theta_{d^\star}\}$, and $P' = P \setminus \{d^\star\}$.

Let $\pi \in \Pi$ be a uniformly good algorithm. By applying Theorem 1 in [9], we know that $\limsup_T R^\pi(T)/\log(T) \geq c(\theta)$, where $c(\theta)$ is the minimal value of the following optimisation problem:

$$
\begin{aligned}
\min \quad & \textstyle\sum_d c_d(\mu_{d^\star} - \mu_d) && (4)\\
\text{s.t.} \quad & \textstyle\inf_{\lambda \in B_d(\theta)} \sum_{l \neq d^\star} c_l I^l(\theta, \lambda) \geq 1, \quad \forall d \in P' && (5)\\
& c_d \geq 0, \quad \forall d. && (6)
\end{aligned}
$$

Now assume that (5) holds. For any $d \in N(d^\star) \cap P'$, for any $\epsilon > 0$, select λ such that $r_d\lambda_d = \mu_{d^\star} + \epsilon$, and for any $l \neq d$, $\lambda_l = \theta_l$. Then $\lambda \in B_d(\theta)$, and hence:

$$\sum_{l \neq d^\star} c_l I^l(\theta, \lambda) = c_d I(\theta_d, \frac{\mu_{d^\star} + \epsilon}{r_d}) \geq 1.$$

We deduce that $c(\theta) \geq c_G^\epsilon(\theta)$ where $c_G^\epsilon(\theta)$ is the minimal value of the optimisation problem:

$$
\begin{aligned}
\min \quad & \textstyle\sum_d c_d(\mu_{d^\star} - \mu_d)\\
\text{s.t.} \quad & c_d I(\theta_d, \tfrac{\mu_{d^\star}+\epsilon}{r_d}) \geq 1, \quad \forall d \in N(d^\star)\\
& c_d \geq 0, \quad \forall d.
\end{aligned}
$$

Hence for any $\epsilon > 0$, $c(\theta) \geq \sum_{d \in N(d^\star)} \frac{\mu_{d^\star} - \mu_d}{I(\theta_d, \frac{\mu_{d^\star}+\epsilon}{r_d})}$. We conclude that $\limsup_T R^\pi(T)/\log(T) \geq c_G(\theta)$. □

PROOF OF THEOREM 3.2

We provide a sketch of proof only due to space limitations, refer to [4] for a complete proof. Let $T > 0$. The regret of $\pi = G - ORS$ up to time T is:

$$R^\pi(T) = \sum_{d \neq d^\star} (\mu_{d^\star} - \mu_d)\mathbb{E}[\sum_{n=1}^{T} \{d(n) = d\}].$$

We decompose the set $\{d(n) = d\}$ into $A_d(n) = \{d(n) = d, L(n) \neq d^\star\}$ (the leader is not $d^\star$) and $B_d(n) = \{d(n) = d, L(n) = d^\star\}$ (the leader is $d^\star$), and analyse the two corresponding contributions to regret. We have:

$$\sum_{d \neq d^\star} (\mu_{d^\star} - \mu_d)\mathbb{E}[\sum_{n=1}^{T} \{A_d(n)\}] \leq r_{d^\star} \sum_{d \neq d^\star} \mathbb{E}[l_d(T)].$$

Now when $L(n) = d^\star$, G-ORS selects a decision $d \in \mathcal{N}(d^\star)$, we deduce that $R^\pi(T)$ is upper bounded by:

$$r_{d^\star} \sum_{d \neq d^\star} \mathbb{E}[l_d(T)] + \sum_{d \in N(d^\star)} (\mu_{d^\star} - \mu_d)\mathbb{E}[\sum_{n=1}^{T} \{B_d(n)\}]$$

The main difficulty consists in bounding the first term, i.e., the average number of times where $d^\star$ is not the leader. The following result provides the required bound. Its proof presented in [4][Theorem C.1] relies on concentration inequalities, and properties of the KL divergence.

$$\mathbb{E}[l_d(T)] = O(\log(\log(T))), \quad \forall d \neq d^\star.$$

From the above theorem, we conclude that the leader is $d^\star$ except for a negligible number of instants (in expectation). When $d^\star$ is the leader, G-ORS behaves as KL-UCB [7] restricted to the set $N(d^\star)$ of possible decisions. Following the same analysis as in [7], we can show that for all $\epsilon > 0$ there are constants C_1 , $C_2(\epsilon)$ and $\beta(\epsilon) > 0$ such that:

$$
\begin{aligned}
&\mathbb{E}[\sum_{n=1}^{T} \{B_d(n)\}] \leq \mathbb{E}[\sum_{n=1}^{T} \{b_d(n) \geq b_{d^\star}(n)\}]\\
&\leq (1+\epsilon)\frac{\log(T)}{I(\theta_d, \frac{r_{d^\star}\theta_{d^\star}}{r_d})} + C_1 \log(\log(T)) + \frac{C_2(\epsilon)}{T^{\beta(\epsilon)}}.
\end{aligned}
$$

Putting pieces together, we get:

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

which concludes the proof. □