Joint Control of Transmission Power and Channel Switching against Adaptive Jamming

Qingsi Wang and Mingyan Liu
Department of EECS
University of Michigan, Ann Arbor

***Abstract*—In this work, we consider the interaction between an adaptive jammer and a user, and we study the joint control of transmission power and channel switching for the user in the presence of jamming interference. Instead of adopting a game-theoretical approach to studying the interaction, which typically requires full knowledge of the jammer about the user and vice versa, we adopt an online learning perspective to model the reasoning of the attacker as well as the defender. We note that the two aspects of the control are typically coupled, and moreover, the tractability of analysis may be limited if the jammer is capable of learning the user's control strategy, thus attacking adaptively. Interestingly, we show that the power control aspect can be in fact decoupled from the channel switching decisions as a result of this interaction, and we develop the explicit form of the optimal control for certain scenarios.**

I. Introduction

In a multi-channel wireless communication system, a user (or node) is typically presented with the challenges of strategically tuning the transmission power and selecting the channel to use. These challenges arise from the energy constraint of the hardware, e.g., a battery-powered device with energy harvesting, and the time-varying nature of the channel quality, including natural fluctuation and interference from other users. In this paper, we consider a particular cause of the variation of the channel condition due to the jamming interference from an attacker, and study the joint control of transmission power and channel switching in this context. We note that the two aspects of the control are typically coupled, and moreover, the tractability of analysis may be limited if the jammer is capable of learning the user's control strategy, thus attacking adaptively.

In this study we are interested in understanding the interaction between an adaptive jammer and a user, and aim to present a framework for analyzing this interaction and studying the joint control of the user. We show that the power control aspect can be decoupled from the channel switching decisions as a result of this interaction, and develop the explicit form of the optimal control for certain scenarios.

Related work. Defense against jamming attacks has been extensively studied in the literature, and a big part focuses on specific attack and defense mechanisms, see e.g., [1], [2] for a collection of jamming attacks and anti-jamming measures. Examples also include using stronger error detection, correction, and spreading codes at the physical layer [3], [4], [5], [6], exploring the vulnerability in the rate adaptation mechanism of IEEE 802.11 [7], and multi-channel jamming using a single cognitive radio [8]. Interestingly, jamming can also be used by legitimate users to achieve physical layer security in the presence of an eavesdropper, see e.g., [9], [10], [11].

The interaction between a jammer and a user/defender is often modeled as a strategic game. Examples include a non-zero-sum game formulation when transmission costs are incurred to both the jammer and the user [12], a random access game [13], a differential game between a mobile jammer and mobile users [14], a Stackelberg game [15], and a zero-sum game framework [16]. Typically the existence of Nash equilibrium strategies is investigated and the equilibrium strategies are identified if they exist under the respective game formulation.

There is also a large body of existing work on the transmission power control problem, especially when the energy-constrained node is capable of harvesting renewable energy. The study of stand-alone optimal power control problems is not the focus of this work, and the interested reader is referred to, e.g., [17], [18], [19], [20] and the references therein.

Our approach. Instead of adopting a game-theoretical approach to studying the interaction between the two sides, which typically requires full knowledge of the jammer

about the user and vice versa, including their respective information/strategy spaces, and infinite rationality to obtain equilibrium solutions, we will simply assume that the jammer is able to learn and adapt over time using its observations of the user's behavior; it need not possess all the information available to the user nor does it presume that the user is rational. To model the jammer's adaptive behavior, we will employ online learning algorithms developed for the class of adversarial or non-stochastic multi-armed bandit problems [21], [22], which provide robust and considerable performance guarantee, without assuming any probabilistic model of the underlying reward process (the user's behavior). We then investigate two cases. In the first case we assume that the user is aware of the type of learning algorithm used by the jammer while in the second case it has no such information and thus must also try to learn.

Main contribution.

- We present an analytical framework of the joint control problem for a class of sublinear regret online learning algorithms that may be adopted by the attacker.
- We show that the optimal transmission power control can be decoupled from the optimal channel switching decisions when channels are symmetric and static, and the overall problem can be reduced to a rate maximization problem with stochastic energy replenishment, to which some existing results may readily apply.
- We also show that given any power control policy that is independent from channel switching decisions, channel selection strategies induced by any pair of sublinear regret learning algorithms can be mutually best responses for the interacting two sides, which resembles game theoretical equilibrium solution concepts.

We proceed as follows. In Section II, we describe the model and formulate the joint control problem. We present the results for the two cases in Sections III and IV, respectively, and Section V concludes the paper.

II. System Model and Problem Formulation

Consider a wireless communication system consisting of a user, a base station (or data sink), and m orthogonal channels denoted as $\mathcal{C} = \{1, 2, \ldots, m\}$. The system operates in discrete time; in each time slot the user attempts to transmit data to the base station. Due to hardware constraints, the user is limited to use at most one or a subset of the channels at any given time, while the base station is capable of receiving data from all channels simultaneously.

The user node is powered by an onboard energy storage device, e.g., battery, with the finite capacity B_{cap}. Let $B(t) \in [0, B_{\text{cap}}]$ be the energy level stored at the node at time t, and we assume the dynamics of the energy level is governed by a sequence of functions $f_t, t = 1, 2, \ldots$, i.e.,

$$B(t+1) = f_t(B(t), E(t), P(t)) \tag{1}$$

with $B(1) = B_{\text{init}}$, where $E(t)$ and $P(t)$ are the energy replenishment and the transmission power at time t, respectively, and B_{init} is the given initial energy level. We assume that the user has a probabilistic model on the energy replenishing process.

There is a jammer/attacker who attempts to interfere with the user's communication. It is capable of transmitting jamming signals over $M < m$ channels simultaneously, within a certain power constraint. Let $\mu : \mathbb{R}_+^2 \to \mathbb{R}_+$ be the rate function. The data rate achievable on a channel is given by $\mu(p, q)$ when the transmission power of the user is p and the interfering power of the jammer is q. We assume that channels are symmetric and static, and μ is thus channel-independent and time-invariant, as suggested by our notation. The rate function is increasing in p and decreasing in q, and we also assume that the attacker jams any channel at a fixed power level q, which is known to the user, e.g. through measurement.

Given any time slot t, we define a variable $Z_k(t)$ such that $Z_k(t) = 0$ if there is no jamming attack in channel k, and $Z_k(t) = 1$ otherwise. In the first case when the user knows the type of algorithm/reasoning the attacker uses, it may regard $Z_k(t)$ as stochastic, i.e., assuming that the attacker behaves probabilistically according to $P(Z_k(t) = 1) = \alpha_k(t)$, though the value of this probability may be unknown to the user. This set of marginal probabilities $\alpha(t) = \{\alpha_k(t), k \in \mathcal{C}\}$ that channel k is jammed at time t, will be called the attacker's adversarial behavior. There are two interpretations of $\alpha(t)$: it can describe a randomized strategy of the attacker, or a probabilistic belief of the user about the likelihood of an attack on any channel. Accordingly, the expected utility U_k that the user derives from using channel k at a given time with the transmission power $P(t)$ is defined as

$$U_k(\alpha_k(t), P(t)) = \mu(P(t), 0)(1 - \alpha_k(t)) + \mu(P(t), q)\alpha_k(t)$$

$$= \mu(P(t),0) - (\mu(P(t),0) - \mu(P(t),q))\alpha_k(t).$$

This is essentially the conditional throughput the user obtains given that it attempts a transmission on this channel when it uses the power $P(t)$ for transmission. In the second case when the user has no such information, it may regard $Z_k(t)$ as a predetermined but unknown number $z_k(t)$. Accordingly, the user's utility, denoted by $\hat{U}_k$, is given by

$$\hat{U}_k(z_k(t), P(t)) = \mu(P(t),0) - z_k(t)(\mu(P(t),0) - \mu(P(t),q)).$$

From the attacker's point of view, it derives utility from the loss of the user due to jamming interference. Given any time slot t, we define a variable $X_k(t)$ such that $X_k(t) = \mu(P(t),0) - \mu(P(t),q)$ if the user transmits on channel k with power $P(t)$, and $X_k(t) = 0$ otherwise. Similar to the user, the attacker can also take on two views of the nature of $X_k(t)$. We will focus on the case that the attacker views $X_k(t)$ as a predetermined but unknown number $x_k(t)$ as we will explain in Section III. The attacker's utility of jamming channel k is given by

$$\hat{U}_k^J(x_k(t)) = x_k(t),$$

Depending on the knowledge of the user on the pattern of adaptive jamming, we then formulate the joint control problem in two cases.

A. Against known adaptive attack

In Section III, we assume the user knows the type of adaptive algorithm used by the jammer, and seeks to make optimal channel switching decisions so as to maximally evade the attack. In particular, the user assumes the attacker behaves probabilistically as the attacker indeed does and knows the value of the adversarial behavior $\alpha(t)$ at the beginning of the time slot t. Note that $\alpha(t)$ can be random itself, with a known probabilistic model to the user in this case, and we will describe the attack pattern in detail in Section III. Thus, the user perceives the channel condition variable $Z_k(t)$ as stochastic. Results obtained in this section are then used as benchmarks in the later section when we examine the more realistic situation where the user does not presume to know the attacker's adaptive behavior.

At each time t, the user decides the control action $A(t) = (P(t), \pi(t))$, where $P(t)$ is the transmission power and $\pi(t)$ is the index of the channel to use. We call the 2-tuple $S(t) = (B(t), \alpha(t))$ the state of the system at time t. We assume that $B(t)$ is perfectly observable to the user, and then so is the state, and we also assume that the user has a perfect recall of all past system states and control actions. We argue later (c.f. the remarks after Theorem 2) that the perfect recall condition on control actions can be significantly weakened for optimal decisions. The user then determines its control action as a function of the history of system states, past control actions, and a private randomization device that is independent from any activity of the attacker. Formally, we have

$$P(t) = g_t^p(A^{[t-1]}, S^{[t]}, \omega^p(t)),$$

and

$$\pi(t) = g_t^\pi(A^{[t-1]}, S^{[t]}, \omega^\pi(t)),$$

where the notation $A^{[t-1]}$ denotes the vector $(A(1), \ldots, A(t-1))$ with $S^{[t]}$ similarly defined, $\omega^p(t)$ (resp. $\omega^\pi(t)$), $t = 1, 2, \ldots$, is a random process with known joint distributions to the user for any finite collection, and g_t^p and g_t^π are the control rules for transmission power and channel switching at time t. Moreover, a power control policy is called feasible if it satisfies the energy causality constraint that $0 \le P(t) \le B(t)$. We denote the control policy – the collection of control rules – by $g^p = (g_t^p, t \ge 1)$ and $g^\pi = (g_t^\pi, t \ge 1)$. We also denote the feasible policy spaces as $\mathcal{G}^p$ and $\mathcal{G}^\pi$.

Given a transmission power profile $P = (P(1), P(2), \ldots)$ and a channel selection sequence $\pi = (\pi(1), \pi(2), \ldots)$ of a pair of policies g^p and g^π, the user collects reward $r(t) = U_{\pi(t)}(\alpha_{\pi(t)}(t), P(t))$ at each time t. The user then considers the following infinite-horizon reward maximization problem:

$$\underset{g^p \in \mathcal{G}^p, g^\pi \in \mathcal{G}^\pi}{\text{maximize}} \quad \overline{r}(g^p, g^\pi) := \liminf_{T\to\infty} \mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T} r(t)\right\}, \quad (2)$$

where the expectation is taken with respect to (w.r.t.) the randomness of system states and the private randomization devices.

B. Against unknown adaptive attack

In Section IV, we consider the more practical scenario where the user has no information on the attack pattern, and it perceives $Z_k(t)$ as a predetermined but unknown number $z_k(t)$. In this case, we call the energy level as the state of the system. We assume the user can observe the value of $z_k(t)$ of the selected channel after transmission at time t, and it has perfect recall of past observations and control actions, as

in the case of a known attack pattern. At each time, the user determines the control action as a function of the history of system states, all past observations and actions, as well as a private randomization device, i.e.,

$$P(t) = \hat{g}_t^p(B^{[t]}, z_\pi^{[t-1]}, A^{[t-1]}, \omega^p(t)),$$

and

$$\pi(t) = \hat{g}_t^\pi(B^{[t]}, z_\pi^{[t-1]}, A^{[t-1]}, \omega^\pi(t)),$$

where $z_\pi^{[t-1]}$ denotes $(z_{\pi(1)}(1), \ldots, z_{\pi(t-1)}(t-1))$. We similarly define the control policies $\hat{g}^p$ and $\hat{g}^\pi$, and the policy spaces $\hat{\mathcal{G}}^p$ and $\hat{\mathcal{G}}^\pi$. The user receives reward $\hat{r}(t) = \hat{U}_{\pi(t)}(z_{\pi(t)}(t), P(t))$ at each time t if it transmits with power $P(t)$, and it consider the same reward maximization problems with the proper policy space and reward function defined above, where the expectation in the objective is taken w.r.t the randomness of the energy replenishing process and the private randomization device.

Note that typically we cannot directly evaluate the objectives in this case, given the unknown and non-stochastic nature of the attack pattern. The optimal control of this type of setting is usually addressed in the framework of non-stochastic online learning, where the existing literature focuses on a "sample-path" argument and seeks the best possible response in terms of minimizing the regret from the performance of ad-hoc strategies for any realization of unknown variables. These learning techniques will become our main model for the adaptive attack throughout the paper and also the countermeasure of the user in this case.

III. Optimal Joint Control Against Known Adaptive Attack

We start by assuming that the user always has data to transmit, though this can be relaxed as described later. The user receives feedback right after a transmission as to whether it has been interfered/jammed. The attacker has to decide which channel to jam and commit to that decision at the beginning of a slot; it however gets to find out the transmission activity on all channels by the end of that time slot. In other words, we assume the attacker needs to make the right decision right at the beginning of a slot in order to have sufficient time for effective jamming. If it chose correctly, it naturally learns the fact that the channel was used by the user and that jamming was successful; however, even if it chose unwisely (one that the user did not use), this does not prevent it from finding out *after the fact* which channel the user actually used for transmission by scanning through the channels.

The attacker is not assumed to know the user's decision making rationale, and thus regards the user activity variable $X_k(t)$ as deterministic but unknown. Given the full information on past user activities on all channels available to the jammer, we assume it adopts an algorithm referred to as Hedge-M shown below. Hedge-M is a multiple-play (attack) extension of the Hedge algorithm, and the latter is a variant introduced by Auer et al. [21] of the original Hedge algorithm developed by Freund and Schapire [23], along the line of work on multiplicative weights learning [24] (see [25] for an in-depth survey and references therein). Hedge-M is reverse-engineered from the algorithm Exp3.M [26], which is a multiple-play algorithm (also an extension of single-play algorithm Exp3 [26]) with partial information (the attacker only observes the activity of the channels it jammed). Hedge-M (or its root Hedge or Exp3) is an online learning algorithm in the adversarial multi-arm bandit setting [21], [22], which presumes no probabilistic behavior pattern of the opponent (in our case, the user). It can be shown to be able to guarantee an order-optimal sublinear weak regret, which in our problem context translates into sublinear "missing" of jamming opportunities compared to always attacking (in hindsight) the most active/used channel under arbitrary user transmission decisions.

Formally, let $x(t) = (x_k(t), \forall k \in \mathcal{C})$, $t = 1, \cdots, T$, over a finite horizon T, and let $\mathcal{C}_M(t)$ denote the set of indices of M channels the attacker jams at time t, which is regarded as a random variable by the user. For any jamming strategy $A = (\mathcal{C}_M(1), \mathcal{C}_M(2), \ldots, \mathcal{C}_M(T))$, the total utility of the attacker is given by

$$G_A(T) = \sum_{t=1}^{T} \sum_{k \in \mathcal{C}_M(t)} \hat{U}_k^J(x_k(t)) = \sum_{t=1}^{T} \sum_{k \in \mathcal{C}_M(t)} x_k(t).$$

while the maximum reward of consistently attacking the M most rewarding channel in hindsight is

$$G_{\max}(T) = \max_{\mathcal{C}_M \subset \mathcal{C}} \sum_{t=1}^{T} \sum_{k \in \mathcal{C}_M} x_k(t),$$

where $\mathcal{C}_M$ is of cardinality M. Hedge-M aims to minimize the gap (i.e., regret) between $G_{\text{Hedge-M}}$ and $G_{\max}$, by selecting channels randomly using a set of marginal probabilities

which is adaptively updated based on past user activities: it always selects the most rewarding channels in retrospect with the highest probability. The algorithm is shown below.

Hedge-M

Parameter: A real number $a > 1$.

Initialization: Set $w_k(1) := 1$ for all $k \in \mathcal{C}$.

Repeat for $t = 1, 2, \ldots, T$

1) If $\max_{k\in\mathcal{C}} \frac{w_k(t)}{\sum_{j=1}^m w_j(t)} > \frac{1}{M}$, then compute $v(t)$ such that
$$\frac{v(t)}{\sum_{k:w_k(t)\geq v(t)} v(t) + \sum_{k:w_k(t)<v(t)} w_k(t)} = \frac{1}{M},$$
and set $\mathcal{C}_0(t) := \{k : w_k(t) \geq v_t\}$. Otherwise, set $\mathcal{C}_0(t) := \emptyset$.
2) Set
$$w'_k(t) = \begin{cases} v(t), & k \in \mathcal{C}_0(t) \\ w_k(t), & k \in \mathcal{C} \setminus \mathcal{C}_0(t) \end{cases}$$
3) Let $\alpha(t) = (\alpha_1(t), \alpha_t(t), \ldots, \alpha_m(t))$ where
$$\alpha_k(t) = M\frac{w'_k(t)}{\sum_{j=1}^m w'_j(t)},$$
and choose M channels with the marginal distribution α, using a subroutine *Dependent Rounding* that returns the set $\mathcal{C}_1(t)$ of channels selected.
4) Observe (reward) vector $(x_1(t), x_2(t), \ldots, x_m(t))$.
5) Set $w_k(t+1) = w_k(t)a^{x_k(t)}$ for all $k \in \mathcal{C}$.

The subroutine Dependent Rounding [27] draws M out of m items with the given marginal distribution, which can be found in the appendix. The performance of Hedge-M is formally characterized by the following theorem, and the proof is omitted for brevity.

Theorem 1: If $a = 1 + \sqrt{2\ln(m/M)/(MT)}$, then $\mathbb{E}G_{\text{Hedge-M}}(T) \geq G_{\max}(T) - \sqrt{2\ln(m/M)MT}$, where the expectation is w.r.t. the randomness in the actions taken by Hedge-M.

We assume the user knows the facts that it is the only user and that the attacker is using Hedge-M; it thus maintains the correct belief about the evolution of the adversarial behavior $\alpha(t)$ as in our general formulation for known attack patterns, where $\alpha(t)$ is determined by Hedge-M, given all actions taken by the user in the past. That is, the evolution of the adversarial behavior can be written as a function the history of all past control actions on channel selection and transmission power:

$$\alpha(t) = h_t(\pi^{[t-1]}, P^{[t-1]}) \tag{3}$$

for some function h_t. Hence, combining with (1) and (3), given any control rules g_t^p and g_t^π, there exist functions $\tilde{g}_t^p$ and $\tilde{g}_t^\pi$ such that

$$P(t) = \tilde{g}_t^p(E^{[t-1]}, \omega^{p,[t]}, \omega^{\pi,[t]}),$$

and

$$\pi(t) = \tilde{g}_t^\pi(E^{[t-1]}, \omega^{p,[t]}, \omega^{\pi,[t]}),$$

where we have suppressed the dependence on the constant initial conditions B_{init} and $\alpha(1)$.

Given any sequence of energy replenishment $E^{[T-1]}$ and outcomes of the private random devices $\omega^{p,[T]}$ and $\omega^{\pi,[T]}$, for any control policies g^p and g^π, let $\boldsymbol{\mu} := (\mu(1), \ldots, \mu(T))$ be the induced rate profile when there is no jamming attack on the transmissions. i.e., $\mu(t) := \mu(P(t), 0)$; similarly, let $\boldsymbol{\mu}^q := (\mu^q(1), \ldots, \mu^q(T))$ where $\mu^q(t) := \mu(P(t), q)$. Let $\overline{\mu}(T) := \frac{1}{T}\sum_{t=1}^T \mu(t)$ and $\overline{\mu}^q(T) := \frac{1}{T}\sum_{t=1}^T \mu^q(t)$. We then have the following result.

Lemma 1: For any polices g^p and g^π,

$$\overline{r}(g^p, g^\pi) \leq \liminf_{T\to\infty} \mathbb{E}\left\{\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)\right\}.$$

Proof: Let $(\mathcal{C}_M(1), \mathcal{C}_M(2), \ldots, \mathcal{C}_M(T))$ be the sequence of M channels selected by Hedge-M, and we have

$$\begin{aligned}
&\mathbb{E}G_{\text{Hedge-M}}(T) \\
&= \mathbb{E}\left\{\sum_{t=1}^T \sum_{k\in\mathcal{C}_M(t)} x_k(t)\right\} = \sum_{t=1}^T\sum_{k=1}^m x_k(t)\alpha_k(t) \\
&= \sum_{t=1}^T(\mu(t) - \mu^q(t))\alpha_{\pi(t)}(t) = \sum_{t=1}^T \mu(t) - \sum_{t=1}^T r(t).
\end{aligned}$$

Note that the above expectation is taken w.r.t. the randomness in Hedge-M. We then obtain

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T r(t) = \overline{\mu}(T) - \frac{1}{T}\mathbb{E}G_{\text{Hedge-M}}(T) \\
&\leq \overline{\mu}(T) - \frac{1}{T}(G_{\max}(T) - \sqrt{2T\ln(m/M)M}) \\
&\leq \frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T) + \sqrt{2\ln(m/M)M/T},
\end{aligned}$$

since $G_{\max}(T) \geq \frac{M}{m}\sum_{t=1}^T(\mu(t) - \mu^q(t))$, and the result then follows. ∎

We note that the quantity $\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)$ can

depend only on the results of power control. Consider then any power control policy $\gamma^p = (\gamma_t^p, t \geq 1)$ that is independent of the past channel switching decisions and adversarial behaviors, i.e.,

$$P(t) = \gamma_t^p(P^{[t-1]}, B^{[t]}, \omega^p(t)).$$

Denote the space of such policies by Γ^p, and note that Γ^p is a subset of $\mathcal{G}^p$. It can be argued that

$$\max_{g^p \in \mathcal{G}^p, g^\pi \in \mathcal{G}^\pi} \liminf_{T\to\infty} \mathbb{E}\left\{\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)\right\} \quad (4)$$

$$= \max_{\gamma^p \in \Gamma^p} \liminf_{T\to\infty} \mathbb{E}\left\{\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)\right\}, \quad (5)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^q$ on the right-hand side of (5) are induced[1] by γ^p given any E^{T-1} and $\omega^{p,[T]}$. Let $\gamma_\star^p$ be an optimal policy for this maximization. Consider also a greedy policy of channel switching, denoted by $\gamma_{\text{greedy}}^\pi$, by which $\pi(t) \in \arg\min_{k\in C} \alpha_k(t)$ for all t. We show that an optimal joint control can be in the form of $\gamma_\star^p$ and $\gamma_{\text{greedy}}^\pi$, by showing that this pair achieves the upper bound established in the right-hand side of (5).

Theorem 2: For any $\gamma^p \in \Gamma^p$ together with $\gamma_{\text{greedy}}^\pi$,

$$\overline{r}(\gamma^p, \gamma_{\text{greedy}}^\pi) = \liminf_{T\to\infty} \mathbb{E}\left\{\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)\right\}.$$

Hence, the optimal transmission power control can be decoupled from the optimal channel switching, and the greedy channel switching policy is optimal given the decoupled power control. This is also the reason why this decoupling is referred to as "one-way".

Proof: Consider the induced rate profiles $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^q$ for any $\gamma^p \in \Gamma^p$. The proof is straightforward by noting that $\min_{k\in C} \alpha_k(t) \leq \frac{M}{m}$. Hence, using the greedy channel switching policy, we have

$$r(t) \geq \mu(t) - (\mu(t) - \mu^q(t))\frac{M}{m} = \frac{m-M}{m}\mu(t) + \frac{M}{m}\mu^q(t),$$

and the result then follows. ∎

Using the greedy policy for channel switching, it is only necessary for the user to have a perfect recall of the last control action and to be able to store and update the value of $G_k(t)$ for all $k \in \mathcal{C}$. We also note that the same result holds even if the user does not always have data to transmit, because an idle slot would not affect the adversarial behavior of the attacker given its full information on the channel activity. The same result can also be extended to the case where the user is able to transmit in multiple channels simultaneously. Furthermore, the above result suggests the joint control problem then reduces to the infinite-horizon problem given by the right-hand side of (5). Note that when a binary collision model is adopted, i.e., $\mu(p,q) = 0$ for any $q > 0$, the joint control problem reduces to the rate maximization problem with the energy causality constraint:

$$\max_{\gamma^p \in \Gamma^p} \overline{\mu}(\gamma^p) := \liminf_{T\to\infty} \mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T} \mu(P(t), 0)\right\}. \quad (6)$$

Some existing results readily apply to (6), see e.g. [19], [20].

IV. Against Unknown Attacks

We now turn to the more realistic case where the user presumes no knowledge of the reasoning used by the attacker, and accordingly employs learning techniques. We show that when the adversary happens to use the online learning techniques like in the previous section, the user can perform optimally as in the last case even without the knowledge of attack pattern. As in the previous section we assume symmetric and static channels.

As before, we will assume that the user can only observe the jamming activity if it occurs in the same channel the user selects. Previously, when the user knows the adaptation the attacker is using, this assumption does not affect the user in any significant way since it is able to completely track the attacker's belief based on its own actions. However, when the user cannot assume knowledge on the attacker's behavior, this assumption implies partial information for the user: it only gets to find out the value z_k if it selects channel k. Consequently we will assume that the user adopts the partial information counterpart of Hedge called Exp3 [21], [22] to update its probability τ_k of choosing channel k in a time slot. The description of Exp3 is given in the appendix; it has a similar sublinear regret like Hedge [21], [22]. Note that Exp3 as a channel switching policy belongs to the policy space $\hat{\mathcal{G}}^\pi$ in our formulation. For the attacker, we will assume it continues to use Hedge-M, though this fact is unknown to the user.

Repeating a similar argument as in the previous section, for any joint control $\hat{g}^p$ and $\hat{g}^\pi$, we can then similarly obtain the same upper bound as the right-hand side of (5):

$$\overline{r}(\hat{g}^p, \hat{g}^\pi) \leq \max_{\gamma^p \in \Gamma^p} \liminf_{T\to\infty} \mathbb{E}\left\{\frac{m-M}{m}\overline{\mu}(T) + \frac{M}{m}\overline{\mu}^q(T)\right\}$$

[1]Using a similar argument, for any γ_t^p, there exists a function $\tilde{\gamma}_t^p$ such that $P(t) = \tilde{\gamma}_t^p(E^{T-1}, \omega^{p,[T]})$

when the attacker uses Hedge-M, where γ^p and Γ^p are defined as in the previous section, and $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^q$ are the induced rate profiles by γ^p given any $E^{[T-1]}$ and $\omega^{p,[T]}$.

On the other hand, consider the joint control given by Exp3 as the channel switching policy with the power control policy $\gamma^p_\star \in \Gamma^p$ that maximizes the right-hand side of (5). Given the induced rate profiles by $\gamma^p_\star$, denoted by $\boldsymbol{\mu}_\star$ and $\boldsymbol{\mu}^q_\star$, define $R_{\max}(T)$ as the user's total reward in retrospect from transmitting on the least jammed channel until a finite horizon T. Using the weak regret result of Exp3 [22, Corollary 3.2], of which the total reward is denoted $R_{\text{Exp3}}(T)$ given $\boldsymbol{\mu}_\star$ and $\boldsymbol{\mu}^q_\star$, we obtain

$$\mathbb{E}R_{\text{Exp3}}(T) \geq R_{\max}(T) - 2.63\sqrt{m\ln(m)T},$$

where the expectation is taken w.r.t. the randomness in Exp3. Since $R_{\max}(T) \geq \sum_{t=1}^T \mu_\star(t) - \frac{M}{m}\sum_{t=1}^T(\mu_\star(t) - \mu^q_\star(t))$, we have

$$\frac{1}{T}\mathbb{E}R_{\text{Exp3}}(T) \geq \frac{m-M}{m}\overline{\mu}_\star(T) + \frac{M}{m}\overline{\mu}^q_\star(T) + o(1), \quad (7)$$

where the $o(1)$ term is w.r.t. the growth of T. Hence, $\overline{r}(\hat{g}^p, \text{Exp3})$ achieves the same upper bound of the average reward established as in the previous sectioin. In other words, Exp3 results in the same average reward for the user compared to the case when it knows that the attacker is using Hedge-M and responds optimally. This shows that there is no loss of optimality when using online learning techniques against an unknown attacker who happens to use a similar algorithm but with more information. Also, the overall joint control problem also reduces to a rate maximization problem with the energy causality constraint.

Interestingly, as we show next, Hedge-M is also the best response for the attacker to the actions of the user in this case, when the user adopts Exp3 as the channel switching policy with $\gamma^p_\star$ for the power control. Recall our notation $A = (\mathcal{C}_M(1), \mathcal{C}_M(2), \ldots, \mathcal{C}_M(T))$ for any jamming strategy and $G_A(T)$ for the total reward of the attacker at T in Section III. Let $\overline{r}^J(A) := \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}G_A(T)$. Since $\frac{1}{T}\mathbb{E}G_A(T) + \frac{1}{T}\mathbb{E}R_{\text{Exp3}}(T) = \overline{\mu}_\star(T)$, where the expectation of G_A is taken w.r.t. to the possible randomness in A, we have

$$\begin{aligned}\overline{r}^J(A) &\leq \liminf_{T\to\infty}\left\{\overline{\mu}_\star(T) - \left(\frac{m-M}{m}\overline{\mu}_\star(T) + \frac{M}{m}\overline{\mu}^q_\star(T)\right)\right\} \\ &= \liminf_{T\to\infty}\frac{M}{m}(\overline{\mu}_\star(T) - \overline{\mu}^q_\star(T)).\end{aligned}$$

Conversely, as we have seen, the average gain of the attacker using Hedge-M is:

$$\frac{1}{T}\mathbb{E}G_{\text{Hedge-M}}(T) \geq \frac{M}{m}(\overline{\mu}_\star(T) - \overline{\mu}^q_\star(T)) + o(1).$$

Hence, Hedge-M and Exp3 are *mutually best responses* given $\gamma^p_\star$. In fact, this claim holds for any γ^p using the same argument as above. Moreover, note also that the above analysis only relies on the regret properties of Hedge-M and Exp3, and their applicability in our context. The mutually best response pair can extend to a much larger family of sublinear regret algorithms that are applicable in the underlying problem.

V. Conclusion

In this paper, we considered the joint control of transmission power and channel selection decisions in a multi-channel system, in the presence of a jamming attacker. We focused on the one-way decoupling of the joint control, and showed that the joint control can be reduced to a (single-channel) rate maximization problem when the attacker uses sublinear-regret online learning algorithms as its jamming strategy. The key to establishing this decoupling is to show that given any joint control policies, the achievable upper bound of the performance metric only depends on the power control decisions in this case. We presented the optimal channel switching policy in two cases, depending whether the user knows the reasoning used by the attacker, and we showed that there is no loss of optimality even if the user has no such knowledge.

References

[1] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The Feasibility of Launching and Detecting Jamming Attacks in Wireless Networks," in *MobiHoc '05*, pp. 46–57, 2005.

[2] A. Wood, J. Stankovic, and G. Zhou, "DEEJAM: Defeating Energy-Efficient Jamming in IEEE 802.15.4-based Wireless Networks," in *SECON '07*, pp. 60 –69, 2007.

[3] G. Noubir and G. Lin, "Low-power DoS Attacks in Data Wireless LANs and Countermeasures," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 7, no. 3, pp. 29–30, 2003.

[4] E. Kehdi and B. Li, "Null Keys: Limiting Malicious Attacks Via Null Space Properties of Network Coding," in *INFOCOM '09*, pp. 1224 –1232, april 2009.

[5] J. Chiang and Y.-C. Hu, "Cross-Layer Jamming Detection and Mitigation in Wireless Broadcast Networks," *Networking, IEEE/ACM Transactions on*, vol. 19, no. 1, pp. 286–298, 2011.

[6] C. Popper, M. Strasser, and S. Capkun, "Anti-jamming Broadcast Communication Using Uncoordinated Spread Spectrum Techniques," *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 5, pp. 703–715, 2010.

[7] G. Noubir, R. Rajaraman, B. Sheng, and B. Thapa, "On the Robustness of IEEE 802.11 Rate Adaptation Algorithms Against Smart Jamming," in *WiSec '11*, WiSec '11, (New York, NY, USA), pp. 97–108, ACM, 2011.

[8] A. Sampath, H. Dai, H. Zheng, and B. Zhao, "Multi-channel Jamming Attacks using Cognitive Radios," in *ICCCN '07*, pp. 352–357, 2007.

[9] R. Negi and S. Goel, "Secret Communication Using Artificial Noise," in *Vehicular Technology Conference*, vol. 3, pp. 1906–1910, 2005.

[10] L. Dong, Z. Han, A. Petropulu, and H. Poor, "Cooperative Jamming for Wireless Physical Layer Security," in *SSP '09*, pp. 417 –420, 31 2009-sept. 3 2009.

[11] S. Gollakota and D. Katabi, "Physical Layer Wireless Security Made Fast and Channel Independent," in *INFOCOM '11*, pp. 1125–1133, 2011.

[12] E. Altman, K. Avrachenkov, and A. Garnaev, "A Jamming Game in Wireless Networks with Transmission Cost," in *Network Control and Optimization*, Springer Berlin Heidelberg, 2007.

[13] Y. Sagduyu and A. Ephremides, "A Game-Theoretic Analysis of Denial of Service Attacks in Wireless Random Access," in *WiOpt '07*, pp. 1 –10, april 2007.

[14] S. Bhattacharya and T. Başar, "Game-theoretic Analysis of an Aerial Jamming Attack on a UAV Communication Network," in *ACC '10*, pp. 818–823, 2010.

[15] V. Navda, A. Bohra, S. Ganguly, and D. Rubenstein, "Using Channel Hopping to Increase 802.11 Resilience to Jamming Attacks," in *INFOCOM '07, Mini-Conference*, pp. 2526–2530, 2007.

[16] K. Pelechrinis, C. Koufogiannakis, and S. Krishnamurthy, "On the Efficacy of Frequency Hopping in Coping with Jamming Attacks in 802.11 Networks," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 10, pp. 3258 –3271, 2010.

[17] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with Energy Harvesting Nodes in Fading Wireless Channels: Optimal Policies," *Selected Areas in Communications, IEEE Journal on*, vol. 29, pp. 1732–1743, Sep. 2011.

[18] A. Sinha and P. Chaporkar, "Optimal Power Allocation for a Renewable Energy Source," in *Communications (NCC), 2012 National Conference on*, pp. 1–5, Feb. 2012.

[19] S. Chen, N. B. Shroff, P. Sinha, and C. Joo, "A Simple Asymptotically Optimal Energy Allocation and Routing Scheme in Rechargeable Sensor Networks," in *INFOCOM'12*, 2012.

[20] Q. Wang and M. Liu, "When Simplicity Meets Optimality: Efficient Transmission Power Control with Stochastic Energy Harvesting," in *INFOCOM'13 mini-conference*, 2013.

[21] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire, "Gambling in a Rigged Casino: The Adversarial Multi-armed Bandit Problem," in *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pp. 322–331, 1995.

[22] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The Nonstochastic Multiarmed Bandit Problem," *SIAM J. Comput.*, vol. 32, no. 1, pp. 48–77, 2003.

[23] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.

[24] N. Littlestone and M. K. Warmuth, "The Weighted Majority Algorithm," *Information and Computation*, vol. 108, no. 2, pp. 212–261, 1994.

[25] S. Arora, E. Hazan, and S. Kale, "The Multiplicative Weights Update Method: a Meta-Algorithm and Applications," *Theory of Computing*, vol. 8, no. 6, pp. 121–164, 2012.

[26] T. Uchiya, A. Nakamura, and M. Kudo, "Algorithms for Adversarial Bandit Problems with Multiple Plays," in *Proceedings of the 21st international conference on Algorithmic learning theory*, pp. 375–389, Springer-Verlag, 2010.

[27] R. Gandhi, S. Khuller, S. Parthasarathy, and A. Srinivasan, "Dependent Rounding and its Applications to Approximation Algorithms," *J. ACM*, vol. 53, no. 3, pp. 324–360, 2006.

APPENDIX A
THE DEPENDENT ROUNDING ALGORITHM

Dependent Rounding

Input: A marginal distribution $(\alpha_k, k \in \mathcal{C})$ and a natural number $M < |\mathcal{C}|$ such that $\sum_{k\in\mathcal{C}} \alpha_k = M$.

Output: A subset $\mathcal{C}_1$ of $\mathcal{C}$ such that $|\mathcal{C}_1| = M$.

Initialization: $p_k = \alpha_k$ for all $k \in \mathcal{C}$.

While $\{k \in \mathcal{C} : 0 < p_k < 1\} \neq \emptyset$ **do**

1) Choose distinct i and j with $0 < p_i < 1$ and $0 < p_j < 1$.
2) Set $a = \min\{1 - p_i, p_j\}$ and $b = \min\{p_i, 1 - p_j\}$.
3) Update p_i and p_j as

$$(p_i, p_j) = \begin{cases} (p_i + a, p_j - a), & \text{w.p. } \frac{b}{a+b} \\ (p_i - b, p_j + b), & \text{w.p. } \frac{a}{a+b} \end{cases}$$

Return $\{k \in \mathcal{C} : p_k = 1\}$.

APPENDIX B
EXP3

Exp3

Parameters: A real $0 < \gamma \leq 1$.

Initialization: Initialize **Hedge**(e).

Repeat for $t = 1, 2, \ldots, T$

1) Get the distribution $p(t) = (p_1(t), p_2(t), \ldots, p_m(t))$ from **Hedge**.
2) Choose action k_t according to the distribution $\hat{p}(t) = (p_1(t), p_2(t), \ldots, p_m(t))$ on channels, where

$$\hat{p}_k(t) = (1 - \gamma)p_k(t) + \frac{\gamma}{m}$$

3) Receive the reward $x_{k_t}(t)$.
4) Feed the simulated reward vector $\hat{x}(t)$ back to **Hedge**, where

$$\hat{x}_k(t) = \begin{cases} \frac{\gamma}{m} \cdot \frac{x_{k_t}(t)}{\hat{p}_{k_t}(t)}, & k = k_t \\ 0, & k \neq k_t \end{cases}$$