Online Interference Mitigation via Learning in Dynamic IoT Environments

Alexandre Marcastel*, E. Veronica Belmega*†, Panayotis Mertikopoulos‡† and Inbar Fijalkow*
* ETIS/ENSEA – UCP – CNRS, Cergy-Pontoise, France
† Inria
‡ French National Center for Scientific Research (CNRS), LIG F-38000 Grenoble, France

***Abstract*—A key challenge for ensuring self-organization capabilities in the Internet of things (IoT) is that wireless devices must be able to adapt to the network's unpredictable dynamics. In the lower layers of network design, this means the deployment of highly adaptive protocols capable of supporting large numbers of wireless "things" via intelligent interference mitigation and online power control. In view of this, we propose an exponential learning policy for throughput maximization in time-varying, dynamic IoT environments where interference must be kept at very low levels. The proposed policy is provably capable of adapting quickly and efficiently to changes in the network and relies only on locally available and strictly causal information. Specifically, if the transmission horizon T of a device is known ahead of time, the algorithm under study matches the performance of the best possible fixed policy in hindsight within an error margin of $O(T^{-1/2})$; otherwise, if the horizon is not known in advance, the algorithm still achieves a $O(T^{-1/2}\log T)$ worst-case margin. In practice, our numerical results show that the interference induced by the connected devices can be mitigated effectively and – more importantly – in a highly adaptive, distributed way.**

***Index Terms*—Arbitrarily time-varying networks, interference mitigation, online optimization, exponential learning**

I. Introduction

The emerging Internet of things (IoT) paradigm is projected to bring together millions – if not *billions* – of disparate wireless "things" (ranging from smartphones and tablets to sensors and wearables), all with widely varying throughput requirements, power characteristics, utilization levels, etc. In this context, two major challenges arise: First, following Moore's prediction on silicon integration, the underlying wireless habitat of IoT networks is expected to exhibit massive device densities, making interference a key limiting factor in achieving a "speed of thought" user experience at the application level [1]. More importantly, the unique mobility attributes of modern wearable devices – coupled with factors such as intermittent user activity and highly variable application demands – introduce an unprecedented degree of temporal variability to IoT networks which can no longer be treated as tame, stationary systems.

Clearly, effective networking in such environments requires the deployment of physical layer protocols that can support large numbers of wireless interfaces via intelligent interference mitigation and distributed power/medium access control [2]. Nevertheless, little progress has been made in designing resource allocation and/or interference mitigation techniques that are capable of operating efficiently and autonomously in dynamic IoT networks that evolve unpredictably over time. On account of this, we focus on the problem of throughput maximization in dynamic IoT environments where interference must be kept at very low levels.

Specifically, the main objective of our paper is to devise highly adaptive and distributed throughput maximization policies that are provably capable of tracking the dynamic evolution of an IoT network while minimizing co-channel interference (CCI) for devices occupying the same wireless band. To do so, we focus on a multi-user IoT composed of several wireless "things", all with different channel and transmission characteristics, and possibly going on-line and off-line at arbitrary times. As a result of device mobility, fading and variable user demands, the network evolves over time in an unpredictable way, so it is not possible to target a fixed operation state; in particular, static solution concepts (such as social optima or Nash equilibria) are no longer relevant.

To circumvent this obstacle, we take an approach based on *no-regret learning* [3], a dynamic optimization paradigm which provides a suitable framework for studying unpredictably varying systems. Building upon these tools, we propose an adaptive power allocation policy inspired from exponential learning [3–5], relying only on strictly causal and local device information. Our first theoretical result shows that if the transmission horizon T is known to the device beforehand, the proposed algorithm matches the performance of the best fixed transmit policy in hindsight within $O(T^{-1/2})$, even though the latter can only be computed with non-causal, future-anticipating capabilities. We further show that this result remains true when the transmission horizon T is not known in advance and the algorithm is used with a variable step-size parameter; in that case, the algorithm's *regret* (the gap between our algorithm and the best fixed policy) grows slightly to $O(T^{-1/2}\log T)$. These results are validated by numerical simulations which show that the network's devices quickly adapt to the variable wireless landscape (in practice, within a few transmission frames), achieving high transmission rates while keeping interference below a fixed threshold.

The closest works to the current paper are [6] and [7]. In [6], the authors used an exponential learning technique to show that the secondary users of a stationary cognitive radio (CR) network converge to a Nash equilibrium solution assuming

This research was supported in part by the Orange Lab Research Chair on IoT within the University of Cergy-Pontoise, the CNRS project REAL.NET–PEPS–JCJC–2016, and by ENSEA, Cergy-Pontoise, France. PM was partially supported by the French National Research Agency under grant no. ANR–13–INFR–004-NETLEARN.

a static environment; however, such static solution concepts are not relevant in highly dynamic IoT environments. In [7], the authors derive a matrix exponential learning policy for throughput maximization in a dynamic, multi-carrier, multiple-input and multiple-output (MIMO) system, but without taking into account the harmful impact of the induced interference. Our work here incorporates an active interference mitigation component in the devices' learning algorithm, thus allowing the system to maintain low interference levels at the device end, despite its inherent temporal variability.

II. System Model and Problem Formulation

Consider a network composed of K mobile devices (smartphones, tablets, wearables and/or other wireless "things") that connect to the Internet via a shared access point (AP). Assuming that the devices communicate over a set $\mathcal{S} = \{1, \dots, S\}$ of S orthogonal channels, the Shannon throughput of the k-th device will be given by the familiar expression

$$R_k(\mathbf{p}; t) = \sum_{s=1}^{S} \log\left(1 + \frac{p_{ks} g_{ks}(t)}{\sigma_s^2(t) + \sum_{j \neq k} p_{js} g_{js}(t)}\right), \tag{1}$$

where p_{ks} is the transmit power of the k-th device over the s-th channel, $\mathbf{p} = (p_{ks})_{k,s}$ is the power profile of all devices in the network, $g_{ks}(t)$ is the time-varying channel gain between the k-th device and the AP, and $\sigma_s^2(t)$ denotes the variance of the channel noise (including thermal, atmospheric and other peripheral interference effects).

Remark II.1. The particularity of this work consists in accounting for the high variability of the network without relying on any assumptions on the evolution of the interference terms, neither on stationarity nor other statistical assumptions.

Given the energy limitations of mobile wireless devices (especially wearable ones), the set of feasible power allocation profiles of the k-th transmitter will be of the form

$$\mathcal{P}_k = \{\mathbf{p}_k \in \mathbb{R}^S : p_{ks} \geq 0 \text{ and } \textstyle\sum_{s=1}^{S} p_{ks} \leq \bar{P}_k\}. \tag{2}$$

In addition to the above, a key challenge in IoT environments is to maintain the interfering power of the aggregate signal at the AP at a reasonably low level, determined by the AP – for instance, in order to allow new devices to connect at any given time, guarantee the QoS requirements of devices with critical roles (such as wireless health and safety equipment), minimize latency due to packet drops and retransmissions, etc. To that end, we assume that the aggregate signal strength at each subcarrier must remain below a given interference threshold I_s, leading to the requirement

$$\sum_{k=1}^{K} p_{ks} g_{ks}(t) \leq I_s \quad \text{for all } s \in \mathcal{S}. \tag{3}$$

The challenge in maintaining the requirement (3) is twofold: First, given that there can be no reliable coordination between devices in a fully decentralized IoT setting, it is not clear how this constraint can be enforced in a distributed, adaptive way (especially when the network's devices do not have perfect channel state information (CSI) at their disposal). Secondly, due to the unpredictable evolution of the devices' connectivity patterns and channel gains $g_{ks}(t)$, a power profile which is admissible at a given transmission frame may fail to satisfy (3) at the subsequent one because of a new device entering the system, the transposition of a harmful scatterer, etc.

To overcome these challenges, instead of treating (3) as a physical constraint at the device end, we posit that each device incurs a virtual penalty when violating it. Specifically, combining this with the Shannon rate function (1), we will focus on the utility model

$$U_k(\mathbf{p}_k; t) = R_k(\mathbf{p}; t) - C_k(\mathbf{p}; t), \tag{4}$$

where the *interference penalty function* $C(\cdot)$ is of the general form

$$C_k(\mathbf{p}; t) = \sum_{s=1}^{S} C\left(\sum_{k=1}^{K} p_{ks} g_{ks}(t) - I_s\right), \tag{5}$$

for some non-decreasing, convex function $C(\cdot)$ which captures the trade-off between higher throughput and the overall multi-user interference (MUI) at the AP. For instance, a standard example of such a penalty is the piecewise linear function

$$C(x) = \begin{cases} \lambda x & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where λ is a positive parameter that controls the balance between high throughput and low interference levels.

In view of the above, we obtain the dynamic optimization problem

$$\begin{aligned} &\text{maximize} && U_k(\mathbf{p}_k; t) \\ &\text{subject to} && \mathbf{p}_k \in \mathcal{P}_k \end{aligned} \tag{P}$$

Given that the objective of each device depends *explicitly* on time (via its dependence on the channel gains $g_{ks}(t)$ and the possibly intermittent connectivity of all other devices in the network), our goal will be to determine a dynamic power allocation policy $\mathbf{p}_k(t)$ that remains as close as possible to the (evolving) solution of (P).

Of course, due to the temporal variability of the channel gains, the power $\mathbf{p}_k^*(t)$ that solves (P) at every given time t cannot be calculated ahead of time with strictly causal information. By this token, we will instead focus on the fixed power allocation profile that is optimal in hindsight, i.e. the solution of the time-averaged problem:

$$\mathbf{p}_k^* \in \arg\max_{\mathbf{p}_k \in \mathcal{P}_k} \sum\nolimits_{t=0}^{T_k} U_k(\mathbf{p}_k; t), \tag{7}$$

where $\mathcal{P}_k$ is the feasible set of the device k defined in (2). As before, the mean optimal solution $\mathbf{p}_k^*$ can only be calculated offline because it requires knowledge of the evolution of the overall system ahead of time, over the entire transmission horizon. As a result, $\mathbf{p}_k^*$ cannot be calculated in practice and only serves as a theoretical benchmark for a dynamic power allocation policy $\mathbf{p}_k(t)$ that relies only on strictly causal information.

To make this comparison precise, we define the (cumulative) *regret* [8, 9] of the k-th device as:

$$\mathrm{Reg}_k(T_k) = \sum\nolimits_{t=1}^{T_k} U_k(\mathbf{p}_k^*; t) - U_k(\mathbf{p}_k; t) \tag{8}$$

Algorithm 1 Exponential learning.

Initialization: $\mathbf{y}_k \leftarrow 0$; $t \leftarrow 0$.

Repeat

$t \leftarrow t+1$;

{ *Pre-transmission phase: set transmit power* }

$p_{ks} \leftarrow \bar{P}_k \frac{\exp(y_{ks})}{1+\sum_{s'=1}^{S}\exp(y_{ks'})}$;

Transmit;

{ *Post-transmission phase: receive feedback* }

estimate $\mathbf{v}_k \leftarrow \partial_{\mathbf{p}_k} U_k(\mathbf{p}_k; t)$;

update scores $\mathbf{y}_k \leftarrow \mathbf{y}_k + \delta_k \mathbf{v}_k$;

until transmission ends.

In words, $\mathrm{Reg}_k(T_k)$ over the transmission horizon T_k measures the cumulative performance gap between the dynamic power strategy $\mathbf{p}_k(t)$ and the optimum profile $\mathbf{p}_k^*$. In particular, if $\mathrm{Reg}_k(T_k)$ grows linearly with T_k, the device is not able to track changes in the system sufficiently fast. Accordingly, we will say that a power allocation policy $\mathbf{p}_k(t)$ leads to *no regret* if

$$\limsup_{T_k\to\infty} \mathrm{Reg}_k(T_k)/T_k \le 0 \quad \text{for all } k, \tag{9}$$

irrespectively of how the system evolves over time. If this is the case, there is no fixed power profile yielding a higher utility in the long run; put differently, (9) provides an asymptotic guarantee that ensures that the policy $\mathbf{p}(t)$ is at least as good as the a posteriori optimal solution of (7).

III. Exponential Learning

To devise an online policy $\mathbf{p}_k(t)$ that leads to no-regret, our main idea will be based on the following two steps: First, each device's policy tracks the direction of gradient (or subgradient) ascent of their utility, without taking into account the problem's requirements as defined in (3). Subsequently, this "aggregated gradient" is mapped back to the feasible region via a suitably chosen exponential map, and the process repeats.

To be more precise, this procedure can be described by the recurrence

$$\begin{aligned} \mathbf{y}_k(t+1) &= \mathbf{y}_k(t) + \delta_k(t)\mathbf{v}_k(t), \\ p_{ks}(t+1) &= \bar{P}_k \frac{\exp\left(\mathbf{y}_{ks}(t+1)\right)}{1+\sum_{s'=1}^{S}\exp\left(\mathbf{y}_{ks'}(t+1)\right)}, \end{aligned} \tag{XL}$$

where $\mathbf{v}_k(t) = \partial_{k_k} U_k(\mathbf{p}_k; t)$ denotes the gradient of the k-th device's utility function and $\delta(t)$ is a non-decreasing step-size parameter (for an algorithmic implementation, see Algorithm 1 above).

Our goal is to examine the no-regret properties of the online power allocation policy (XL). To do so, let V_k denote an upper bound for $\mathbf{v}_k$, i.e.

$$\|\mathbf{v}_k\|^2 \le V_k^2. \tag{10}$$

For instance, if $C(\cdot)$ is defined as in (5) and (6), the bound can be computed explicitly, giving

$$V_k^2 = S\,\bar{g}_k^2\left(\lambda^2 + 1/\sigma^4\right), \tag{11}$$

where $\sigma^2 = \min_{s,t}\{\sigma_s^2(t)\}$ and $\bar{g}_k = \max_{s,t}\{g_{ks}(t)\}$.

With all this at hand, our first result (see the appendix for a sketch of the proof) concerns the case where the transmission horizon is known in advance (for instance, as in a timed call), and (XL) is employed with a constant, optimized step-size:

Theorem 1. *Assume that the online policy* (XL) *is run for a given time horizon T_k with the optimized step-size $\delta_k^* = V_k^{-1}\sqrt{\log(1+S)/T_k}$. Then, it enjoys the regret bound*

$$\mathrm{Reg}_k(T_k) \le 2V_k\bar{P}_k\sqrt{T_k\log(1+S)}. \tag{12}$$

Consequently, the devices' average regret $\mathrm{Reg}_k(T_k)/T_k$ *vanishes as $O(T_k^{-1/2})$, i.e.* (XL) *leads to no regret.*

The above result relies on the devices knowing their own transmission horizon T_k in advance. If this is not the case, it is more advantageous to consider a strictly decreasing step-size so as to reduce the algorithm's jitter in fluctuations of unknown length. We illustrate this in Theorem 2 below (proven in the appendix):

Theorem 2. *Assume that the online policy* (XL) *is run for an unknown time horizon T_k with the variable step-size $\delta_k(t) = a_k t^{-1/2}$ for some $a_k > 0$. Then, it enjoys the regret bound:*

$$\mathrm{Reg}_k(T_k) \le \bar{P}_k\left(\frac{\log(1+S)}{a_k} + a_k V_k^2\right)T_k^{1/2} + a_k\bar{P}_k V_k^2 T_k^{1/2}\log T_k. \tag{13}$$

Consequently, the devices' average regret $\mathrm{Reg}_k(T_k)/T_k$ *vanishes as $O(T_k^{-1/2}\log T_k)$, i.e.* (XL) *leads to no regret.*

This implies that, in both cases (known vs. unknown horizon), the rate of regret minimization depends on the system parameters. We further remark that the devices' average regret vanishes faster if the transmission horizon T_k is known in advance, but the resulting logarithmic disparity ($\log T_k$) is fairly moderate. This disparity can be overcome completely by means of a more complicated step-size policy known as a "doubling trick" [9] but, for simplicity, we do not present this approach here.

IV. Numerical results

To validate our theoretical results in a realistic environment, we focus on a heterogeneous IoT network composed of a large number of mobile devices that are going on-line and off-line in a random way. This wireless system operates over a 10 MHz band centered around the carrier frequency $f_c = 2$ GHz and divided into 512 subcarriers spaced at 19.5 kHz. We further consider a variable number of mobile devices ranging from 50 to 200, positioned inside a square cell of side 2 km following a Poisson point process, and which share the entire band. The interference tolerance in (3) is the same for all the subcarriers ($I_s = -110\,\mathrm{dBm}$ for all s) and the maximum power of each device varies from $0.5W$ to $2W$. For algorithmic purposes, we use the variable step-size $\delta_k(t) = \delta/\sqrt{t}$ for all k, and we assume that the duration of each communication frame is 5 ms.

The channels between the wireless devices and the AP are generated according to the widely used COST-HATA model

[10] with fast and shadow-fading attributes as in [11]. Each mobile device speed chosen arbitrarily between 0 km/h and 130 km/h so as to account for a wide spectrum of wireless devices (smartphones, wearables, etc.).

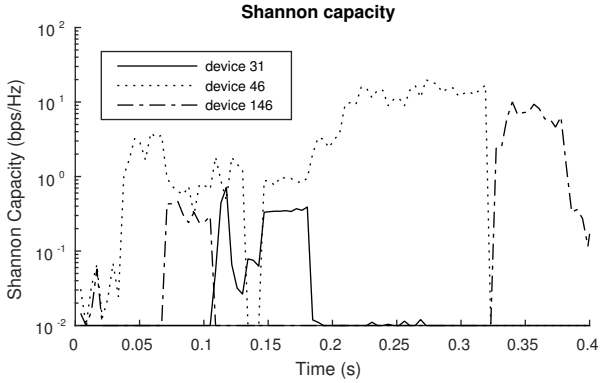

Fig. 1. Evolution on Shannon capacity as function of time. The significant falls result from the penalty applied when the interference requirement (3) is not met. When the rate curves are interrupted, it means that the corresponding device is off-line.

In Fig. 1, we plot the Shannon capacity for three randomly chosen devices and for $\delta = 0.1$ and λ chosen so that $\lambda I_s = 100$. The significant falls in throughput result from the penalty imposed to the devices' utility function when the interference requirement (3) is violated; also, when a rate curve is interrupted, it means that the corresponding device has disconnected from the system.

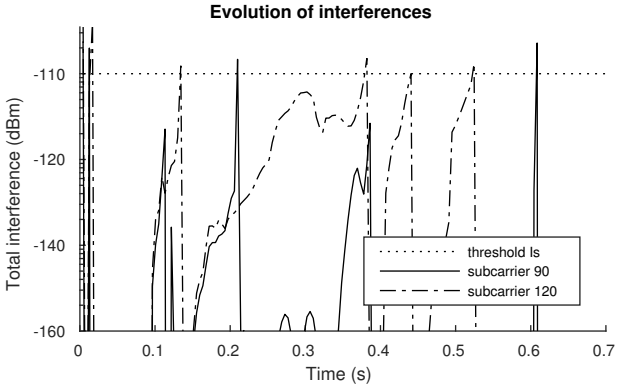

Fig. 2. Overall interference as function of time. When the interference constraints are violated, the penalty results in a drastic reduction in the devices' transmit powers.

In Fig. 2, we plot the evolution of the overall interference in two randomly chosen subcarriers and with step-size and penalty parameters as above. We notice that, in the beginning of the algorithm, the interference constraints are violated but after a few iterations the interference level falls under the maximum threshold as a result of the penalty term. The same can be observed whenever the interference constraints are violated due to the system's variability.

In Fig. 3, we plot the evolution of the devices' average regret as a function of time for different system loads: 50, 100 and 200

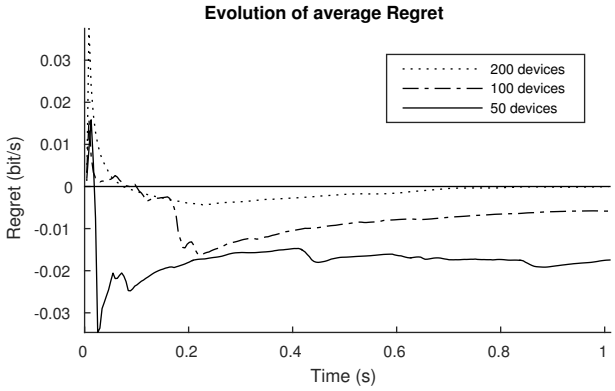

Fig. 3. Evolution of the devices' average regret as function of time for different system loads. The online power allocation policy quickly leads to zero average regret.

connecting devices and for $\delta = 10$, $\lambda I_s = 100$. We see that the average mobile device's regret goes to zero relatively quickly depending on the number of devices. Hence, the online power allocation policy we propose matches the best fixed transmit profile within a few number of iterations, despite the channels' significant variability over time. We also remark that the higher the number of overall devices in the system, the faster the average regret goes to zero. Intuitively, when the number of devices grows large, some randomness is lost as the overall interference term becomes more and more deterministic.

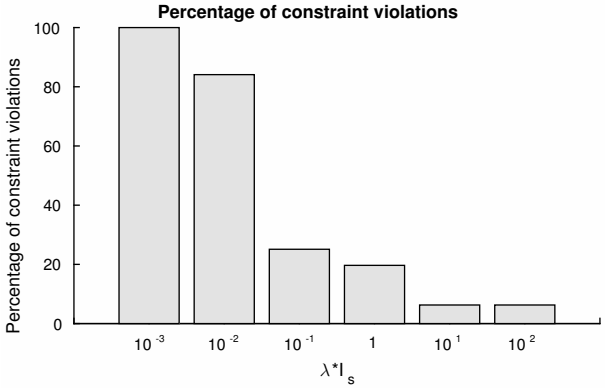

Fig. 4. Fraction of time at which the devices violate the interference constraints. The higher is λ, the higher is the penalty in case of interference constraint violations. This results in less violations by the mobile devices.

Finally, in Fig. 4, we plot the fraction of time at which the mobile devices violate the overall interference requirement in at least one channel for $\delta = 100$. As expected, higher values of λ lead to fewer violation of the requirement (3). Therefore, the exponential learning policy (DXL) with the cost function defined in (5) and (6) allows for an efficient use of the total bandwidth by limiting the interference despite the unpredictability of the system's variation over time.

V. Conclusions

In this paper, we have investigated a dynamic IoT network that varies arbitrarily with time and in which a large number of

devices interfere with each other. We show that the proposed exponential learning algorithm allows the devices to adapt their power allocation policies to these dynamic changes in an optimal way regarding the tradeoff between throughput and harmful interference they inflict. Moreover, our simulations show that the overall network interference can be controlled efficiently and in a distributed way despite the arbitrary and unpredictable network variations. Our approach is based on online convex optimization and learning and makes a clean break from existing cellular resource allocation and interference management techniques that often rely on various stationarity assumptions on the channels' distribution.

Appendix

To prove Theorems 1 and 2, consider the potential function $f(\mathbf{y}) = \bar{P}\log\left(1+\sum_{s'=1}^{S}\exp(y_{s'})\right)$. A brief calculation then shows that (XL) can be written equivalently as

$$\begin{aligned}\mathbf{y}_k(t+1) &= \mathbf{y}_k(t) + \delta(t)\mathbf{v}_k(t),\\ \mathbf{p}_k(t+1) &= \nabla f_{\mathbf{y}_k}(\mathbf{y}_k(t+1)).\end{aligned} \tag{A.14}$$

With this in mind (and dropping the device index k for simplicity of presentation), we have:

Proof of Theorem 1: The first step in proving (13) is to use the concavity of the devices' utility function to write

$$\mathrm{Reg}(T) \leq \frac{1}{\delta}\langle \mathbf{y}(T)|\mathbf{p}^*\rangle - \sum\nolimits_{t=1}^{T}\langle \mathbf{v}(\mathbf{p}(t);t)|\mathbf{p}(t)\rangle, \tag{A.15}$$

where we have used the fact that $\mathbf{y}(t+1) = \mathbf{y}(t) + \delta\,\mathbf{v}(\mathbf{p}(t))$ and that $\mathbf{v}(\mathbf{p}(t)) = \partial_{\mathbf{p}}U(\mathbf{p}(t))$ by construction. A second-order Taylor estimate now yields

$$f(\mathbf{y}(t+1) \leq f(\mathbf{y}(t)) + \delta\left\langle \mathbf{v}(\mathbf{p}(t);t)\middle|\nabla_{\mathbf{y}} f(\mathbf{y}(t))\right\rangle + \delta^2\bar{P}V^2, \tag{A.16}$$

where $V = \max_{\mathbf{p}}\|\mathbf{v}(\mathbf{p})\|$. Hence, by (A.14), we get:

$$\mathrm{Reg}(T) \leq \frac{1}{\delta}\langle \mathbf{y}(T)|\mathbf{p}^*\rangle + \frac{1}{\delta}\left[f(0) - f(\mathbf{y}(T))\right] + \delta\bar{P}V^2T. \tag{A.17}$$

A standard duality argument then yields

$$\mathrm{Reg}(T) \leq \frac{1}{\delta}\left[f^*(\mathbf{p}^*) + \bar{P}\log(1+S)\right] + \delta\bar{P}V^2T, \tag{A.18}$$

where $f^*(\mathbf{p}) = \sum_s p_s\log p_s + (1-P)\log(1-P)$ is the Legendre transform of f. With $f^*(\mathbf{p}) \leq 0$ for all $\mathbf{p}\in\mathcal{P}$, we conclude that

$$\mathrm{Reg}(T) \leq \frac{\bar{P}\log(1+S)}{\delta} + \delta\bar{P}V^2T, \tag{A.19}$$

Optimizing the RHS of (A.19) with respect to δ yields the optimal step-size $\delta = V^{-1}\sqrt{\log(1+S)/T}$ and (12) follows. ∎

Proof of Theorem 2: To bound the regret under a variable step-size δ, it will be convenient to consider the weighted regret

$$\overline{\mathrm{Reg}}(T) = \sum_{t=1}^{T}\delta(t)\left(U(\mathbf{p}^*;t) - U(\mathbf{p}(t);t)\right), \tag{A.20}$$

where $\delta(t)$ is the variable step-size sequence used in (XL). Then, as in the proof of Theorem 1, concavity yields:

$$\overline{\mathrm{Reg}}(T) \leq \langle \mathbf{y}(T)|\mathbf{p}^*\rangle - \sum_{t=1}^{T}\delta(t)\langle \mathbf{v}(\mathbf{p}(t);t)|\mathbf{p}(t)\rangle. \tag{A.21}$$

Thus, arguing as in the proof of Theorem 1, we get

$$\overline{\mathrm{Reg}}(T) \leq \langle \mathbf{y}(T)|\mathbf{p}^*\rangle + \left[f(0) - f(\mathbf{y}(T))\right] + \bar{P}V^2\sum_{t=1}^{T}\delta^2(t), \tag{A.22}$$

and hence:

$$\overline{\mathrm{Reg}}(T) \leq \bar{P}\log(1+S) + \bar{P}V^2\sum_{t=1}^{T}\delta^2(t), \tag{A.23}$$

where we used the fact that $f^*(\mathbf{p}) \leq 0$ for all $\mathbf{p}\in\mathcal{P}$.

To bound the device's unweighted regret $\mathrm{Reg}(t)$, we will resort to a summability criterion of Hardy [14] which allows us to compare weighted sums – in our case, $\mathrm{Reg}(T)$ and $\overline{\mathrm{Reg}}(T)$. In particular, note that the step-size sequence $\delta(t) = at^{-1/2}$ satisfies *a)* $\delta(t) \leq \delta(t+1)$; and *b)* $\sum_{t=1}^{T}\delta(t)/\delta(T) = O(T)$. Therefore, by Theorem 14 in [14], we get

$$\begin{aligned}\frac{1}{T}\mathrm{Reg}(T) \sim \frac{\overline{\mathrm{Reg}}(T)}{\sum_{t=1}^{T}\delta(t)} &\leq \frac{\bar{P}\log(1+S) + \bar{P}V_k^2\sum_{t=1}^{T}\delta^2(t)}{\sum_{t=1}^{T}\delta(t)}\\ &\leq \frac{\bar{P}\log(1+S)}{a\sqrt{T}} + \frac{\bar{P}V_k^2a(1+\log T)}{\sqrt{T}},\end{aligned} \tag{A.24}$$

and the bound (13) follows. ∎

References

[1] Huawei Technologies, "5G: A technology vision," White paper, 2013.

[2] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's INTRAnet of things to a future INTERnet of things: A wireless- and mobility-related view," *IEEE Wireless Commun.*, vol. 17, no. 6, pp. 44–51, December 2010.

[3] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[4] P. Mertikopoulos and A. L. Moustakas, "Learning in the presence of noise," in *GameNets '09: Proceedings of the 1st International Conference on Game Theory for Networks*, 2009.

[5] P. Mertikopoulos and W. H. Sandholm, "Learning in games via reinforcement and regularization," *Mathematics of Operations Research*, 2016.

[6] S. D'Oro, P. Mertikopoulos, A. L. Moustakas, and S. Palazzo, "Interference-based pricing for opportunistic multi-carrier cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6536–6549, December 2015.

[7] P. Mertikopoulos and E. V. Belmega, "Transmit without regrets: online optimization in MIMO–OFDM cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 1987–1999, November 2014.

[8] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[9] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[10] COST Action 231, "Digital mobile radio towards future generation systems," European Commission, final report, 1999.

[11] G. Calcev, D. Chizhik, B. Göransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 2, pp. 389–403, 2007.

[12] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, "Regularization techniques for learning with matrices," *The Journal of Machine Learning Research*, vol. 13, pp. 1865–1890, 2012.

[13] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.

[14] G. H. Hardy, *Divergent Series*. Oxford University Press, 1949.