

Blind Optimal User Association in Small-Cell Networks

Livia Elena Chatzieftheriou
Athens University of Economics
and Business, Greece
liviachatzi@aueb.gr

Apostolos Destounis
Huawei Technologies, France
Apostolos.Destounis@huawei.com

Georgios Paschos
Amazon, Luxembourg
paschosg@amazon.com

Iordanis Koutsopoulos
Athens University of Economics
and Business, Greece
jordan@aueb.gr

Abstract—We learn optimal user association policies for traffic from different locations to Access Points (APs), in the presence of unknown dynamic traffic demand. We aim at minimizing a broad family of α -fair cost functions that express various objectives in load assignment in the wireless downlink, such as total load or total delay minimization. Finding an optimal user association policy in dynamic environments is challenging because traffic demand fluctuations over time are non-stationary and difficult to characterize statistically, which obstructs the computation of cost-efficient associations. Assuming arbitrary traffic patterns over time, we formulate the problem of online learning of optimal user association policies using the Online Convex Optimization (OCO) framework. We introduce a periodic benchmark for OCO problems that generalizes state-of-the-art benchmarks. We exploit inherent properties of the online user association problem and propose PerOnE, a simple online learning scheme that dynamically adapts the association policy to arbitrary traffic demand variations. We compare PerOnE against our periodic benchmark and prove that it enjoys the no-regret property, with additional sublinear dependence of the network size. To the best of our knowledge, this is the first work that introduces a periodic benchmark for OCO problems and a no-regret algorithm for the online user association problem. Our theoretical findings are validated through results on a real-trace dataset.

I. INTRODUCTION

Communication networks in the Beyond 5G (B5G)/6G era are envisioned to support ultra-low latency and bandwidth-demanding services, like those enabled by Internet of Things (IoT) or autonomous vehicles. Two key technological enablers of such services in future communication networks are novel network architectures and the embedded use of Artificial Intelligence (AI) [1]. The new architectures will generalize the Coordinated MultiPoint transmission (CoMP), where APs cooperate to jointly serve requests within their coverage area, and each user's traffic may be served by more than one AP. The pervasive introduction of AI at the network edge, including distributed algorithms for proactive learning and prediction of unknown dynamic processes in the system, will enable the self-optimization of network resource allocation.

In the envisioned ultra-dense wireless networks, devices will be in range of multiple Access Points (APs). These enhanced association possibilities will bring more degrees of freedom, and additional possibilities for optimization. The numerous devices and association alternatives call for a fast and agile

This work was supported by the CHIST-ERA LeadingEdge project, call on "Smart Distribution of Computing in Dynamic Networks" (SDCDN).

user-to-AP association scheme. This is of vital importance for the upcoming bandwidth-demanding services, especially for the downlink, that supports the majority of traffic. Moreover, traffic demand at different locations heavily fluctuates during the day. This could happen, for example, due to sudden changes in the existing sources, or due to new unpredictable sources of traffic. Thus traffic is generally non-stationary during the day, which complicates its accurate statistical characterization and precludes the use of approaches that operate under stationary regimes, such as Lyapunov optimization.

In this work we perform blind user associations on-the-fly, without any assumption or information about the actual traffic demand. We use the Online Convex Optimization (OCO) framework to produce updated solutions incrementally, by readjusting existing ones as new samples are observed. We consider arbitrarily time-varying traffic demand for different locations and allocate it to APs, which we model as queues that capture their own load. These queues form an association policy whose cost belongs to a broad family of α -fair functions of the load at APs, including as special cases several objectives, such as delay or load minimization. We aim at producing association policies that minimize regret, *i.e.*, the deviation of the cost of our online association, from that of the optimal offline association that knows in hindsight the traffic variations. We introduce OPS, a novel periodic benchmark that generalizes state-of-the-art, and then propose PerOnE: an online algorithm that quickly adapts to unpredictable traffic variations, and that learns scalable and asymptotically optimal user association policies for downlink traffic routing to different locations.

A. Contributions

The contributions of our work to the literature are as follows:

- We provide a model and an OCO formulation for the problem of Online Learning (OL) of how to dynamically associate traffic of geographic locations (and therefore users) to APs. Our objective cost function models various targets for communication networks, such as AP load or delay minimization.
- We introduce Optimal Periodic Static (OPS), a novel periodic benchmark for OL problems that generalizes state-of-the-art. In cases of traffic periodicity, a benchmark where the association is the same during the day is not

suitable. OPS is appropriate to compare against, because the optimal policy will most likely be periodic as well.

- We identify and exploit inherent properties of the online problem and design PerOnE, an efficient online algorithm that produces cost-effective association policies under arbitrary changes in traffic demand, with lack of information about the actual generated traffic and its statistical properties. PerOnE stems from Online Mirror Descend.
- We prove PerOnE’s asymptotical optimality, as it achieves regret sublinear to the time-horizon against OPS, that knows traffic variations in hindsight. Further, PerOnE’s regret also scales sublinearly with the network size, which renders it a valid association scheme for the upcoming large wireless networks.
- Our evaluation with publicly available traffic traces confirms the derived analytical results, showing that our algorithm achieves zero regret asymptotically. In fact, its performance appears to be near-optimal with respect to a dynamic algorithm that chooses the optimum user association in each time slot.

In section II we present the state-of-the-art. In section III we describe the model and the static user association problem. In section IV we introduce and analyse OPS. In section V we perform a transformation of the static formulation concluding to an OCO formulation. We then design PerOnE, proving that its regret against OPS is sublinear both to the time horizon and to the problem dimension. Finally, in section VI we evaluate our scheme on a real traffic dataset.

II. RELATED WORK

User association (UA). A widely adopted optimization framework is Network Utility Maximization (NUM) [2], which is exemplified further for AP association. It considers a broad family of convex utility functions of the APs’ load, capturing a variety of objectives, such as load balancing. The following works also consider convex cost functions. In [3] an iterative, distributed and deterministic UA policy that is asymptotically optimal for NUM is presented. The authors in [4] propose an exponentiated gradient algorithm for NUM, proving its convergence rate to the optimum UA. In [5] load balancing across APs is considered. Iterative and combinatorial algorithms that perform local adjustments are presented. In [6] the dynamic load balancing is studied by capturing the system state with fluid equations, and an asymptotically optimal simple myopic strategy is presented. The authors in [7] predict future traffic based on the traffic history by using robust optimization tools and propose an iterative UA technique that minimizes costs.

UA is seen jointly with channel assignment in [8] for minimizing the number of channels needed to serve users. After applying an iterative load balancing algorithm, the problem reduces to a simple channel allocation problem. The work [9] additionally considers transmission power, quantifying limits of the achievable gains. In [10] and [11] UA is seen jointly with content caching for cache hit ratio maximization, and low-complexity practical schemes are presented. The fast-converging scheme of [10] iterates between UA and content

caching, while in [11] users are initially clustered based on their content preferences, and then clusters are assigned to APs. The work [12] additionally considers content recommendation. A simple three-step scheme that sequentially performs a preference-aware UA with service guarantees, a recommendation-aware cache placement, and an adjustment of content recommendations, reveals the gains that can be achieved when UA is considered jointly with content caching and recommendations (as introduced in [13]). Despite their interesting results, works [2]–[5], [8]–[12] consider only static UA instances, work [6] focuses on load balancing, and work [7] performs complex computations on the historical traffic.

OCO theory. The goal in OCO is the minimization of regret against a static benchmark, where regret is the worst-case deviation of the performance of online algorithms from the optimal algorithm that knows all data in hindsight, but is restricted to a single action for the entire time horizon T . The following works consider convex and Lipschitz-continuous objective functions, adversarial constraints and decisions taken over a convex set. In [14] a general class of Online Gradient Ascent (OGA) algorithms with $O(\sqrt{T})$ regret is introduced. The authors in [15] substitute OGA’s projection with a Frank-Wolfe linear optimization step, achieving $O(\sqrt{T})$ regret for stochastic and adversarial costs. In [16] time-varying stochastic constraints under a stochastic Slater assumption are studied, and a drift-plus-penalty algorithm with $O(\sqrt{T})$ expected regret is presented. In [17] regret is systematically balanced with constraint violation. Combining stochastic optimization [16] and standard OCO [14] methods, $O(KT/V + \sqrt{T})$ regret for $O(\sqrt{VT})$ constraint violation is achieved, where $K = T^k$, $k \in [0, 1)$ and $V \in [K, T)$. These works do not consider the dimension of the problem in their solutions, which in our case is the size of the network, and either consider no constraints [14], or rely on heavier assumptions on the input [16], [17].

OCO in network resource allocation. The authors in [18] study online content caching under unknown file popularity. Their no-regret algorithm adapts caching and routing decisions to any file request pattern. In [19] an asymptotically optimal online learning algorithm for video rate adaptation in HTTP Adaptive Streaming under no channel model assumptions is presented. The work [20] studies network power and bandwidth allocation under adversarial costs with bounded variations in consecutive slots. Constraints are satisfied on average, tolerating instantaneous violations. Under an additional Slater assumption, their algorithm achieves sublinear regret against a benchmark that takes the optimal decision in each time slot.

Our work is the first one that applies OCO to the minimum-cost UA problem. Our scheme achieves no-regret in UA decisions, with sublinear dependence both on the time horizon and on the network size, under no assumptions on the input. Our work also introduces a novel periodic benchmark that generalizes state-of-the-art.

III. SYSTEM MODEL AND PROBLEM FORMULATION

Basic definitions. We start by providing some definitions and function properties that are needed throughout the paper.

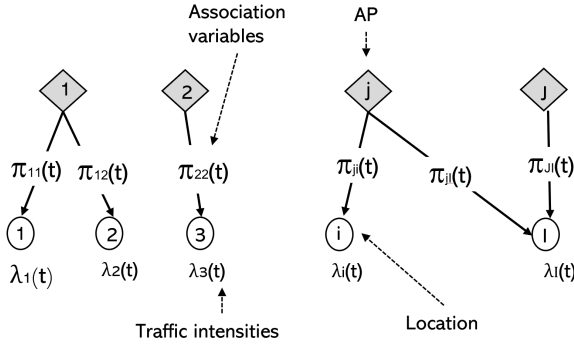


Fig. 1. At each time slot t location $i \in \mathcal{I}$ requests traffic with intensity $\lambda_i(t)$, which can be split into portions $\pi_{ji}(t)\lambda_i$ and served by different APs $j \in \mathcal{N}^i$ in its neighbourhood.

Although we later consider differentiable cost functions, the results of this paper are valid for any other cost function, considering $\nabla f(\mathbf{x})$ to also stand for a subgradient of $f(\cdot)$ at point \mathbf{x} .

- **Convexity.** A function $f(\mathbf{x}) : A \rightarrow B$ is convex iff $\forall \mathbf{x}_1, \mathbf{x}_2 \in A, f(\mathbf{x}_1) - f(\mathbf{x}_2) \leq \langle \nabla f(\mathbf{x}_1), \mathbf{x}_1 - \mathbf{x}_2 \rangle$, for $\langle \mathbf{a}, \mathbf{b} \rangle$ the inner product of \mathbf{a} and \mathbf{b} . If it exists, the Hessian matrix of a convex function is positive semi-definite, and vice versa.

- **p -norm and its dual norm.** Let $\mathbf{x} \in \mathbb{R}^d$. Its p -norm is defined as $\|\mathbf{x}\|_p := \left(\sum_{i=1}^d |x_i|^p \right)^{1/p}$. A q -norm is said to be the dual of p -norm iff $\frac{1}{p} + \frac{1}{q} = 1$.

- **Lipschitz-continuity.** A function $f(\mathbf{x}) : A \rightarrow B$ is Lipschitz-continuous iff the p -norm of the gradient is bounded, i.e., if $\exists L : \|\nabla f(\mathbf{x})\|_p = L < +\infty$.

- **Strong convexity.** A function $f(\mathbf{x}) : A \rightarrow B$ is σ -strongly-convex w.r.t. a p -norm iff $\forall \mathbf{x}_1, \mathbf{x}_2 \in A, f(\mathbf{x}_1) \geq f(\mathbf{x}_2) + \langle \nabla f(\mathbf{x}_1), \mathbf{x}_2 - \mathbf{x}_1 \rangle + \frac{\sigma}{2} \|\mathbf{x}_2 - \mathbf{x}_1\|_p$.

Model components. We consider downlink transmissions in a geographical area that is partitioned into locations $i \in \mathcal{I}$ and is covered by a set \mathcal{J} of APs. We define the "neighbourhood" set $\mathcal{N}_j, j \in \mathcal{J}$, as the subset of locations that can be served by AP j . Similarly, $\mathcal{N}^i, i \in \mathcal{I}$, is the subset of APs that can serve traffic of location i . An overview of our model and relevant notation are given in Fig. 1 and Table I, respectively.

Location traffic. We denote as $\boldsymbol{\lambda} = (\lambda_i)_{i \in \mathcal{I}}$ the traffic intensity vector. Each element $\lambda_i \geq 0$ is the aggregate amount (in packets/second) of requested traffic of all users in location i , modeled as a random variable from a general distribution.

Access Point (AP) load. The traffic requested by a location i can be served by multiple APs, those in \mathcal{N}^i . An association policy determines the association control variables $\pi_{ji} \in [0, 1]$ denoting the fraction of traffic λ_i which is routed from AP j to location i . Each location's demand must be entirely served, so its association variables are constrained to lie in the probability simplex. Thus, $\forall i \in \mathcal{I}, (\pi_{ji})_{j \in \mathcal{J}} \in \Pi$, where:

$$\Pi = \left\{ \mathbf{x} \in [0, 1]^{|\mathcal{J}|} : \sum_{j \in \mathcal{J}} x_j = 1 \right\}. \quad (1)$$

Following an association decision $\boldsymbol{\pi} = (\pi_{ji})_{j \in \mathcal{J}, i \in \mathcal{I}}$, AP j transmits an aggregate demand intensity $\sum_{i \in \mathcal{N}_j} \lambda_i \pi_{ji}$. The packet transmission process at each AP is modeled as a queuing process. Prior work [3] has shown that statistical

Model			
\mathcal{I}	Set of locations i	\mathcal{J}	Set of APs j
\mathcal{N}^i	Set of neighbour APs for location i	\mathcal{N}_j	Set of neighbour locations for AP j
λ_i	Intensity of traffic requested at i	π_{ji}	Fraction of λ_i routed to i by j
ρ_j	Total load at AP j	ρ_0	Load threshold
Π	Probability simplex	Ω	Feasibility set
$\phi_\alpha(\cdot)$	Cost function	T	Time horizon
t	Time slot	K	Number of zones in each period
\mathcal{W}_k	Time window: Set of slots in zone k		

Equivalent problem formulation and Association Algorithm

$V(\cdot)$	Penalty-featured costs	Ω'	Extended feasibility set
L	Lipschitz constant for V	$h(\cdot)$	Regularization function
$g(\cdot)$	Mirror function	Θ	Matrix with gradient information
t_k^t	τ -th time slot in window \mathcal{W}_k		

TABLE I
NOTATION TABLE

multiplexing effects can be captured by modeling this queue with processor sharing service. Assuming the packets have exponentially distributed sizes with mean $1/\omega$, and denoting as C_{ji} the average transmission rate from BS j to location i (averaged over the channel statistics), the load ρ_j of BS j is

$$\rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{i \in \mathcal{N}_j} \frac{\lambda_i \pi_{ji}}{\omega C_{ji}}.$$

Let $\boldsymbol{\rho} = (\rho_j)_{j \in \mathcal{J}}$. The AP traffic load is a measure of the percentage of time the AP is busy with packet transmission. When $\rho_j < 1$, AP j is stable in the sense that its packet transmission queue does not grow unbounded. Values close to 1 indicate large delays. If $\rho_j > 1$, the AP queue is unstable and grows without limit. It results in infinite delays and bad user experience, and therefore must be avoided. To ensure stability and a high-quality service in terms of delay for the end-users, association decisions $\boldsymbol{\pi}$ are constrained so that:

$$\rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda}) \leq \rho_0, \forall j \in \mathcal{J}, \quad (2)$$

where $\rho_0 \in (0, 1)$ a load threshold. Combining (1) and (2), the feasible set for association variables is:

$$\Omega = \left\{ \boldsymbol{\pi} \in \Pi^{|\mathcal{I}|} : \sum_{i \in \mathcal{N}_j} \frac{\lambda_i \pi_{ji}}{\omega C_{ji}} \leq \rho_0, \forall j \in \mathcal{J} \right\}. \quad (3)$$

Cost function. Let $\phi(\boldsymbol{\pi}, \boldsymbol{\lambda})$ be the system cost as a result of association policy $\boldsymbol{\pi}$ under traffic $\boldsymbol{\lambda}$. Our cost functions belong to the following family of convex and Lipschitz-continuous in $\boldsymbol{\pi}$ functions [2], for $\boldsymbol{\rho} \leq \boldsymbol{\rho}_0$ and $\alpha \geq 0$:

$$\phi_\alpha(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{j \in \mathcal{J}} \phi_\alpha^j(\boldsymbol{\pi}, \boldsymbol{\lambda}), \quad (4)$$

$$\text{where } \phi_\alpha^j(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \begin{cases} \frac{1}{\alpha-1} (1 - \rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda}))^{1-\alpha}, & \alpha \neq 1 \\ -\log(1 - \rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda})), & \alpha = 1 \end{cases}. \quad (5)$$

To confirm convexity in $\boldsymbol{\pi}$ when $\boldsymbol{\rho} < \boldsymbol{\rho}_0$, observe that the cost functions are twice differentiable with positive second derivative, hence their Hessian matrix is positive semidefinite, which implies convexity. To confirm Lipschitz-continuity in $\boldsymbol{\pi}$, observe that when $\rho_j < \rho_0$, then $\forall j, i, \nabla \phi_\alpha(\pi_{ji}, \lambda_{ji})$ is bounded, and so is the p -norm $\|\nabla \phi_\alpha(\boldsymbol{\pi}, \boldsymbol{\lambda})\|_p$. We will rely on both the convexity and the Lipschitz-continuity of the cost function to design our online user association algorithm and prove its performance guarantees.

Different values of α lead to different cost functions. For example, for $\alpha = 0$, (4) reduces to the total system load, $\phi_0(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{j \in \mathcal{J}} \rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda})$. For $\alpha = 2$, it is $\phi_2(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{j \in \mathcal{J}} (1 - \rho_j(\boldsymbol{\pi}, \boldsymbol{\lambda}))^{-1}$, and (4) is equivalent to the average delay experienced by a typical demand flow in a stationary

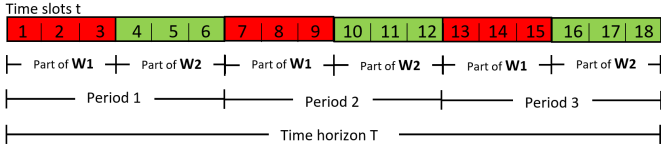


Fig. 2. Toy example to demonstrate periodicity, with $T = 18$ time slots in the time horizon, $P = 3$ periods, $K = 2$ time zones in each period, $Z = 3$ time slots in each time zone during each period. The time windows are: $\mathcal{W}_1 = \{1, 2, 3, 7, 8, 9, 13, 14, 15\}$, $\mathcal{W}_2 = \{4, 5, 6, 10, 11, 12, 16, 17, 18\}$.

system under a temporal fair scheduler, *e.g.*, round robin [3].

Optimal user association for known demand. If the traffic demand vector λ is known, the association policy that minimizes the system costs is found by solving problem:

Problem 1 (Optimal user association for known demand).

$$\begin{aligned} & \min_{\pi} \phi_{\alpha}(\pi, \lambda), \\ \text{s.t.} \quad & \sum_{j \in \mathcal{N}^i} \pi_{ji} = 1, \quad \forall i \in \mathcal{I}, \\ & \rho_j(\pi, \lambda) \leq \rho_0, \quad \forall j \in \mathcal{J}. \end{aligned}$$

Problem 1 is a convex minimization problem of a Lipschitz-continuous cost function, on the intersection of simplex and hyperplane constraints. Therefore, it can be solved through convex optimization methods [21]. Such instances are already studied in literature. In this work we focus on online instances, where the traffic demand is unknown at the time of the decision.

Time dynamics. We capture time dynamics by denoting as $\lambda(t)$, $\pi(t)$, and $\phi_{\alpha}(\pi(t), \lambda(t))$, the traffic vector, association decision, and resulting system cost during time slot $t \in \{1, 2, \dots, T\}$, where T a time horizon.

IV. A NOVEL PERIODIC BENCHMARK AND REGRET

Adversarial Online Learning. In realistic conditions, the traffic $\lambda(t)$ for the next time slot is unknown. Therefore, the association decisions $\pi(t)$ for slot t will be computed based on knowledge of traffic demands $\lambda(t-1)$. After the decision $\pi(t)$ is taken, the actual demand $\lambda(t)$ emerges, and the actual value $\phi_{\alpha}(\pi(t), \lambda(t))$ of the cost function is revealed. This lack of information during the decision $\pi(t)$ at time slot t may imply additional costs, or even instability of AP packet transmission queues, due violation of the load threshold.

An appropriate setting for such online optimization problems is Online Convex Optimization (OCO) [22], [23]. We assume traffic vectors $\lambda(t)$, and therefore costs $\phi_{\alpha}(\pi(t), \lambda(t))$, to be arbitrarily selected by an adversary who tweaks them without adhering to any probability distribution, aiming at obstructing our decisions. While in reality the traffic vectors are not changed by such an adversary, this framework offers a convenient way to design algorithms with provable worst-case guarantees under arbitrary variations of system parameters.

Traffic periodicity. The inherent nature of human activity results in traffic periodicity. For example, daily and weekly patterns can be observed due to people going at work or returning at home. Motivated by this we introduce our periodic benchmark. It generalizes state-of-the-art and characterizes the

Algorithm 1 Optimal Periodic Static (OPS) benchmark policy

Input: Traffic vectors $\lambda(t)$, $t \in \{1, \dots, T\}$, partition of time horizon in time windows \mathcal{W}_k , $k \in \{1, \dots, K\}$.

Output: K optimal periodic static policies $\pi^* = \{\pi^*[k]\}_k$, one for each time window \mathcal{W}_k .

- 1: **for** $k = 1$ to K **do**
- 2: Compute optimal static association policy in zone k ,

$$\pi^*[k] = \arg \min_{\pi \in \Omega} \sum_{t \in \mathcal{W}_k} \phi_{\alpha}(\pi, \lambda(t)) \quad (6)$$

- 3: **end for**

performance of online algorithms, being in between of the two extremes: the static benchmark [14] and the dynamic one.

We divide the time horizon T in P periods, and each period p in K time zones. Without loss of generality, let each time zone k contain Z time slots in each period. Then $P = T/KZ$, and each period p contains KZ slots. This naturally defines a time window \mathcal{W}_k for each time zone k , which includes all time slots that belong to time zone k , across all periods. It is:

$$\mathcal{W}_k = \left\{ t : t = KZ(p-1) + Z(k-1) + \tau, \text{ where} \right.$$

$$\left. \tau \in \{1, \dots, L\}, p \in \{1, \dots, P\} \right\}, k = 1, \dots, K.$$

The K time zones define the manner in which each period is partitioned, while time windows include all time slots of a time zone across the time horizon. This partitioning of the time horizon captures any type of periodicity, *e.g.*, daily, weekly, or any underlying combination. Assuming daily periodicity in our toy example of Fig. 2, the time horizon T is divided in $P = 3$ days, each having $K = 2$ zones. During each period, each zone contains $Z = 3$ time slots, each of 4 hours duration.

We want to stress that *we do not consider periodicity on the traffic demands*: traffic vectors λ are considered to have *arbitrary variations* during time. On the contrary, we aim to capture possible (approximate) periodic-like “patterns” or “trends” that may exist. In fact, a key contribution of this work is the introduction of the following periodic benchmark.

Regret against the Optimal Periodic Static (OPS) algorithm: a novel periodic benchmark. Given a sequence of traffic vectors $\lambda(1), \lambda(2), \dots, \lambda(T)$ over a time horizon T , OPS consists in finding K user association policies $\pi^*[1], \pi^*[2], \dots, \pi^*[K]$, one for each time zone k . Each static policy $\pi^*[k]$ is optimal regarding only traffic loads in the respective time window \mathcal{W}_k , and is defined as in (6). This novel periodic benchmark exploits possible approximate traffic periodicity and allows the comparison of dynamic online policies to static association rules that change according to the general traffic characteristics in each window. For example, it is possible to consider two different association policies, one for peak hours and one for hours with low traffic, and compare our dynamic policy against these. In the toy example of fig. 2, OPS would find two static association policies: one for \mathcal{W}_1 and one for \mathcal{W}_2 . We provide its pseudocode in Algorithm 1.

A performance metric that characterizes the learning performance of an online algorithm is *regret*: the difference between the performance, which in our case is the experienced cost, between an online policy and a benchmark. Let $\pi_A(t)$ be the decision taken by an online algorithm A at slot t . The regret

$Reg_A(T, K)$ of A with respect to OPS, for K time zones in each period over a time horizon T , is:

$$Reg_A(T, K) := \sum_{t=1}^T \phi_a(\pi_A(t), \lambda(t)) - \sum_{k=1}^K \sum_{t \in \mathcal{W}_k} \phi_a(\pi^*[k], \lambda(t)). \quad (7)$$

OPS vs. existing benchmarks and regrets. We remind the reader of the optimal static and optimal dynamic benchmark policies as defined in [14], which we denote as π_S^* and $\pi_D^*(t)$, respectively. The optimal static benchmark knows all traffic changes in hindsight and finds *one* user association policy π_S^* that minimizes the costs over the entire time horizon, *i.e.*,

$$\pi_S^* := \arg \min_{\pi \in \Omega} \sum_{t=1}^T \phi_a(\pi, \lambda(t)). \quad (8)$$

On the contrary, the optimal dynamic benchmark knows all traffic changes but aims at minimizing the cost functions *for each time slot t* and finds one user association $\pi_D^*(t)$:

$$\pi_D^*(t) := \arg \min_{\pi \in \Omega} \phi_a(\pi, \lambda(t)), \forall t = 1, \dots, T. \quad (9)$$

OPS generalizes the state-of-the-art benchmarks, and the respective regrets against them. It is easily verifiable that:

- For $K = 1$, then $\mathcal{W}_k = \{1, \dots, T\}$, $\pi^*[k] \equiv \pi_S^*$, and (7) reduces to the static regret.
- For $K = T$, then $\mathcal{W}_k = \{t = k\}$, $\pi^*[k] \equiv \pi_D^*(t)$, and (7) reduces to the dynamic regret.

Online learning with “no regret”. A desirable property for the regret is to scale sublinearly with the time horizon T , *i.e.*, $Reg_A(T, K) = o(T)$. In this case,

$$\lim_{T \rightarrow +\infty} \frac{Reg_A(T, K)}{T} = 0,$$

and the online algorithm A is said to have “no regret”, which means that it learns to perform as well as the benchmark asymptotically as the time horizon $T \rightarrow +\infty$.

Another desirable feature for online algorithms is to have scalable regret. This happens when regret is also sublinear to the problem dimension d , which in our case equals $|\mathcal{J}| \cdot |\mathcal{I}|$. It means that by increasing the size of the network by a unit, a sublinear increase in the regret is implied. At the moment, most regret results arrive at \sqrt{dT} . In low dimensions, this is a good result, implying the ability to learn quickly. However, as d starts to grow and becomes $d \sim T$, the above expression results in a regret $O(T)$, which means that learning is not attainable in the long run. Indeed, large systems may require a very large horizon T to learn - unless we are able to decrease the dependence of regret expression to d . The above is thus a property of vital importance for the envisioned future large-scaled communication networks.

V. ONLINE USER ASSOCIATION ALGORITHM WITH NO REGRET

A. Augmented penalty function

In order to avoid overloading cells, we reformulate the user association problem with the use of a penalty function that is added to the total cost, while removing the constraints. The penalty is active and adds to the cost when constraints are violated, *i.e.*, when $\exists j : \rho_j > \rho_0$. The set of optimal solutions remains the same, because the structure of the problem and the coupling with the load-constraints now appear in the objective.

Algorithm 2 Online Mirror Descent (OMD)

Input: Mirror function $g : \mathbb{R}^{|\mathcal{J}||\mathcal{I}|} \rightarrow \Omega'$, stepsize η , objective function $V(\cdot)$

Output: User association $\pi(t), \forall t = 1 \dots T$

- 1: **Initialize:** $\Theta(1) = \mathbf{0}$
 - 2: **for** $t=1, 2, \dots, T$ **do**
 - 3: decide association $\pi(t) = g(\Theta(t))$
 - 4: update $\Theta(t+1) = \Theta(t) - \nabla V(\pi(t), \lambda(t))$
 - 5: **end for**
-

Our penalty function $B_j(\pi(t), \lambda(t))$ for overloading AP j could be any convex and Lipschitz-continuous in ρ function, such as $B_j(\pi(t), \lambda(t)) = \psi \nabla \phi_\alpha^j(\rho_0) \cdot (\rho_j - \rho_0)$, where $\psi > 0$ a penalty factor for AP-overloading. This captures the cost for each overloaded AP as the linear extension of the cost function at the overloading point ρ_0 . Then, the cost function becomes:

$$V(\pi(t), \lambda(t)) = \sum_{j \in \mathcal{J}} V_j(\pi(t), \lambda(t)), \text{ where} \quad (10)$$

$$V_j(\pi(t), \lambda(t)) = \begin{cases} \phi_\alpha^j(\rho_j), & \rho_j \leq \rho_0 \\ \phi_\alpha^j(\rho_0) + \nabla \phi_\alpha^j(\rho_0) \cdot (\rho_j - \rho_0), & \rho_j > \rho_0 \end{cases}.$$

The optimal user association problem reduces to:

Problem 2 (Online user association for unknown demand).

$$\min_{\pi(t) \in \Omega'} \sum_{t=1}^T V(\pi(t), \lambda(t)) - \sum_{k=1}^K \sum_{t \in \mathcal{W}_k} V(\pi^*[k], \lambda(t)).$$

where

$$\Omega' = \left\{ \pi : \pi \in \Pi^{|\mathcal{I}|} \right\}. \quad (11)$$

This is a typical formulation for online learning problems. It aims at finding online a sequence of association policies $\pi(t)$, $t = 1, \dots, T$ that minimize regret, *i.e.*, the deviation of online decisions from those of an offline benchmark, which in this work is OPS. Compared to Problem 1, the feasibility set is expanded to a simplex for each location i and it remains convex. This penalty formulation will enable us to perform a customised modification of a traditional algorithm, based on the specific characteristics of the new feasible space. The objective function remains convex and Lipschitz-continuous, as the sum of such functions. Both convexity and Lipschitz-continuity are crucial properties for proving that the online algorithm we will design has no regret against OPS.

We will analyze the regret with respect to this augmented cost. Since the linear part comes into play only when $\rho_j > \rho_0$ (which does not happen in the benchmark), a sublinear regret here implies sublinear regret for the $\phi_\alpha(\cdot)$ functions as well.

B. PerOnE: Online user association with no regret

Online Mirror Descent. A general class of online schemes with no regret against the static benchmark is Online Mirror Descent (OMD) [22], presented in Algorithm 2. It gives the opportunity to exploit the feasibility set of our problem, and leads to decision updates that lie in the feasible set without the need for expensive projections. OMD computes the current decision from the previous one using a simple gradient update rule. Let $\Theta(t)$ be a matrix of dimension $|\mathcal{J}||\mathcal{I}|$, initialized¹

¹Since the first available traffic vector is $\lambda(1)$, the first update in (12) cannot be performed for $t < 2$. Thus, the initialization is performed for $t = 1$, instead of the common choice $t = 0$.

as $\Theta(1) = \mathbf{0}$, and updated as

$$\Theta(t) = \Theta(t-1) - \nabla V(\boldsymbol{\pi}(t-1), \boldsymbol{\lambda}(t-1)), t > 1. \quad (12)$$

During slot t it is given as input to a "link" function $g(\cdot)$, that combines it with the previous decision $\boldsymbol{\pi}(t-1)$ and "mirrors" it to a feasible association decision $\boldsymbol{\pi}(t)$. More specifically, the updated user association is $\boldsymbol{\pi}(t) = g(\Theta(t))$, where

$$g(\Theta) := \arg \min_{\boldsymbol{\pi} \in \Omega'} \{h(\boldsymbol{\pi}) - \langle \eta \Theta, \boldsymbol{\pi} \rangle\}, \quad (13)$$

with η a stepsize, and $h(\cdot)$ a "regularization" function that is strongly-convex with respect to a norm over the feasible set Ω' , where Ω' as in (11).

Regularization function. The regularization function ensures stability of the decision and, if chosen appropriately, it leads to solutions that exploit the geometry of the problem, do not need expensive projections to the feasible space, and enjoy the no-regret property.

In our setting, we aim at finding associations that lie in the unit simplex for each location. Thus, each association policy is basically a set of probability distributions, one for each location. Since the feasibility set regarding location i is the probability simplex, the most natural regularization function would be the Gibbs-Shannon entropy, $h_i(\boldsymbol{\pi}) = \sum_{j \in \mathcal{J}} \pi_{ji} \log \pi_{ji}$, which would give the well known Exponentiated Gradient Descend (EGD). Here we consider the regularization function

$$h(\boldsymbol{\pi}) := \sum_{i \in \mathcal{I}} h_i(\boldsymbol{\pi}) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \pi_{ji} \log \pi_{ji}, \quad (14)$$

which for a given user association policy equals the aggregate entropy of the associations for all locations. In Appendix A we prove that:

Lemma 1. *The modified entropic regularization function in (14) is $|\mathcal{I}|$ -strongly convex w.r.t the 1-norm.*

Normalized exponentiated gradient. Combining (14) and (13), we get

$$g(\Theta) = \arg \min_{\boldsymbol{\pi} \in \Omega'} \left\{ \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \pi_{ji} \log \pi_{ji}, - \langle \eta \Theta, \boldsymbol{\pi} \rangle \right\}.$$

By differentiating with respect to π_{ji} , we get:

$$\frac{\partial g(\Theta)}{\partial \pi_{ji}} = \log(\pi_{ji}) - \eta \Theta_{ji} + 1,$$

where Θ_{ji} is the element of matrix Θ related to AP j and location i . This becomes zero at $\pi_{ji} = e^{\eta \Theta_{ji} - 1}$. In order to ensure that the updated association variables π_{ji} will lie in the unit simplex for each location i , we need to normalize the association of each location. Each element Θ_{ji} is thus "mirrored" through the exponentiated mirror function to:

$$g_{ji}(\Theta) = \frac{e^{\eta \Theta_{ji}}}{\sum_{j \in \mathcal{J}} e^{\eta \Theta_{ji}}}. \quad (15)$$

Each association variable π_{ji} is then updated through this mirroring as:

$$\pi_{ji}(t+1) = g_{ji}(\Theta(t+1)) \stackrel{(15)}{=} \frac{e^{\eta \Theta_{ji}(t+1)}}{\sum_{j \in \mathcal{J}} e^{\eta \Theta_{ji}(t+1)}}$$

$$\stackrel{(12)}{=} \frac{e^{\eta \Theta_{ji}(t)} e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t), \boldsymbol{\lambda}_i(t))}}{\sum_{j \in \mathcal{J}} e^{\eta \Theta_{ji}(t)} e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t), \boldsymbol{\lambda}_i(t))}} \cdot \frac{\sum_{j \in \mathcal{J}} e^{\eta \Theta_{ji}(t)}}{\sum_{j \in \mathcal{J}} e^{\eta \Theta_{ji}(t)}} \stackrel{(15)}{=} \frac{\pi_{ji}(t) e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t), \boldsymbol{\lambda}_i(t))}}{\sum_{j \in \mathcal{J}} \pi_{ji}(t) e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t), \boldsymbol{\lambda}_i(t))}}. \quad (16)$$

Algorithm 3 Periodic Online Exponentiated (PerOnE)

Input: Set of locations \mathcal{I} , APs \mathcal{J} and neighbouring APs $\mathcal{N}^i, \forall i$, penalty-featured cost functions $V(\cdot)$, partition of time horizon T in time windows $\mathcal{W}_k, k = 1, \dots, K$, step size η .

Output: User association $\boldsymbol{\pi}(t), \forall t = 1, \dots, T$

- 1: **for** $t = 1, 2, \dots, T$ **do**
 - 2: Identify time window $\mathcal{W}_k \ni t$
 - 3: **if** $t = t_k^1$ for \mathcal{W}_k **then**
 - 4: Initialize association as $\pi_{ji}(t_k^1) = \begin{cases} \frac{1}{|\mathcal{N}^i|} & j \in \mathcal{N}^i \\ 0, & j \notin \mathcal{N}^i \end{cases}$ (17)
 - 5: **else if** $t = t_k^\tau$ for \mathcal{W}_k **then**
 - 6: Update association as
 - 7:
$$\pi_{ji}(t_k^{\tau+1}) = \frac{\pi_{ji}(t_k^\tau) \cdot e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t_k^\tau), \boldsymbol{\lambda}_i(t_k^\tau))}}{\sum_{j \in \mathcal{J}} \pi_{ji}(t_k^\tau) \cdot e^{-\eta \nabla V(\boldsymbol{\pi}_{ji}(t_k^\tau), \boldsymbol{\lambda}_i(t_k^\tau))}} \quad (18)$$
 - 7: **end if**
 - 8: Observe actual traffic $\boldsymbol{\lambda}(t)$
 - 9: Compute gradient $\nabla V(\boldsymbol{\pi}(t), \boldsymbol{\lambda}(t))$
 - 10: **end for**
-

This mapping is a simple normalization of the product of the previous association, multiplied with a negative exponentiation of the gradient of the objective function in the previous step. The controller, thus, needs only the value of $\nabla V(\cdot)$ in order to decide the association of all locations i to their neighbourhood APs $j \in \mathcal{N}^i$. Using an adequate (for the geometry of the problem) normalization function leads us to decision updates that are always on the feasible set, avoiding expensive projections that would be otherwise necessary but prohibitive for large-scale networks.

PerOnE: Our PERiodic, ONline, Exponentiated gradient association algorithm with "no regret". We design it based on the normalized exponentiated gradient-based association update (16). We refer to it as PerOnE and we provide its pseudocode in Algorithm 3. PerOnE exploits possible traffic periodicity and operates in each time window $\mathcal{W}_k, k = 1, \dots, K$ separately.

Let t_k^1, t_k^τ and $t_k^{|\mathcal{W}_k|}$ be the first, the τ -th, and last time slot in time window \mathcal{W}_k , respectively. For the first slot t_k^1 in each window \mathcal{W}_k , PerOnE does not have a previous user association $\boldsymbol{\pi}$ to rely on, nor any prior information about traffic vectors $\boldsymbol{\lambda}(t)$ for $t \in \mathcal{W}_k$. Therefore, it simply splits the requested traffic evenly across neighbouring APs. At time $t = t_k^{\tau+1}$, it updates the association variables $\pi_{ji}(t_k^{\tau+1})$ as in (16). For this update, it is based on the slot $t = t_k^\tau$, which precedes $t_k^{\tau+1}$ within the same time window $\mathcal{W}_k, \forall k$, as shown in (18).

PerOnE, is a simple, projection-free and cost-efficient modification of the OMD. It also achieves a sublinear bound on the regret against the OPS benchmark over the time-horizon T and over the total number $|\mathcal{J}||\mathcal{I}|$ of decision variables. Let

$$M_I = \max_j |\mathcal{N}_j| \quad \text{and} \quad M_J = \max_i |\mathcal{N}^i| \quad (19)$$

be the maximum number of locations that are in range of an AP j in the system, and the maximum number of APs that a location i is in range of, respectively. Then it holds:

Theorem 1 (PerOnE No-regret). *For a Lipschitz-continuous and convex objective function $V(\cdot)$, with L Lipschitz constant, and M_I, M_J as in (19), a stepsize η and T time horizon,*

Scheme	$\rho_0 = 1$			$\rho_0 = 0.5$		
	$K = 24$	$K = 12$	$K = 2$	$K = 24$	$K = 12$	$K = 2$
OPS	234	936	2808	4212	3744	2808
PerOnE	127	57	9	145	72	11
Optimum (OPS, $K = T$)	0			0		
PerOnE ($K = 1$)	1			3		

TABLE II

AGGREGATE CONSTRAINT VIOLATIONS DURING TIME HORIZON, FOR DIFFERENT VALUES OF ZONES K AND LOAD THRESHOLDS ρ_0 .

PerOnE achieves the regret bound:

$$\text{Reg}(T, K) \leq \frac{KM_I \log(M_J)}{\eta M_J} + \frac{\eta TL^2}{2|\mathcal{I}|}.$$

In particular, for stepsize $\eta = \sqrt{\frac{2KM_I|\mathcal{I}|\log(M_J)}{TL^2M_J}}$, and since $M_I \leq |\mathcal{I}|$ and $\log(M_J) \leq M_J$, we get:

$$\text{Reg}(T, K) \leq \sqrt{\frac{2KM_I TL^2 \log(M_J)}{|\mathcal{I}|M_J}} \leq L\sqrt{2KT}.$$

Please refer to Appendix B for a proof.

Remark 1. The EGD, obtained as the OMD with regularization function the entropic $h_i(\pi)$ for only one location, has a regret of \sqrt{T} on the horizon, and a $\log(|\mathcal{J}||\mathcal{I}|)$ dependence on the number of association variables [22].

Remark 2. Our entropic function $h(\pi)$ that considers multiple locations, and the initialization step in (17), imply a dependence of regret on topological characteristics such as the maximum number of locations M_I that an AP has in its range, and the maximum number M_J of APs in whose range the location belongs. Overall, its regret is sublinear on the total number of association decision variables $|\mathcal{J}||\mathcal{I}|$. Moreover, in realistic systems, the impact of the linear dependence in M_I and that of the logarithmic dependence in M_J , on regret is very limited. In fact, their values can be considered constant compared to the system's dimension $|\mathcal{J}||\mathcal{I}|$, due to the progressively decreasing range of APs as technology evolves, which results in \mathcal{N}^i and \mathcal{N}_j being progressively smaller sets.

PerOnE's regret follows the \sqrt{T} dependence of EGD, and it also depends on the number K of time zones. For $K = o(T)$, the regret is sublinear to the time horizon T , i.e., $\lim_{T \rightarrow +\infty} \frac{\text{Reg}(T, K)}{T} = 0$, which means that PerOnE learns association policies that are asymptotically optimal. The standard [14] static regret is obtained for $K = 1$, and aligns with the above. For $K = O(T)$, the regret scales linearly with time, which is aligned with the impossibility result stated in [24]: when the adversary can change its decision in each time slot, no-regret is not-attainable without other assumptions on the input. Indeed, PerOnE will just play the initialization for each t , or any other linear scaling, since it will play very few rounds for each part of the horizon. The state-of-the-art dynamic regret, obtained for $K = T$, aligns with the above.

VI. NUMERICAL EVALUATION

A. System architecture and traffic demand

We perform our evaluation on the internet traffic activity of the publicly available dataset [25]. It provides the demand of

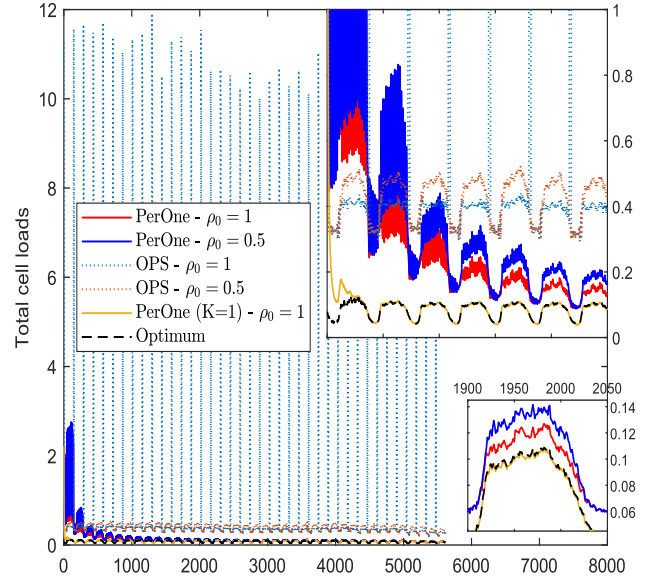


Fig. 3. Total cell loads vs. time, for $\alpha = 0$ and cost function $\phi_a(\rho) = \sum_{j \in \mathcal{J}} \rho_j$ and for load threshold $\rho_0 = 1$ and $\rho_0 = 0.5$, and for $K = 24$.

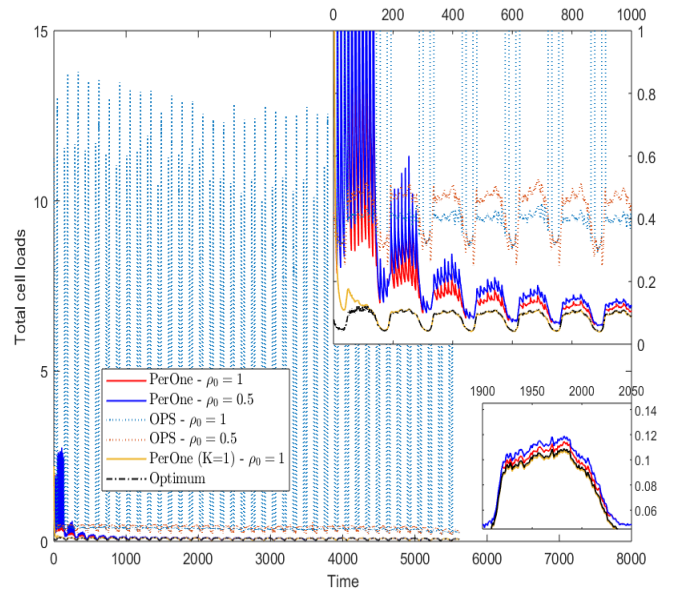


Fig. 4. Total cell loads vs. time, for $\alpha = 0$ and cost function $\phi_a(\rho) = \sum_{j \in \mathcal{J}} \rho_j$ and for load threshold $\rho_0 = 1$ and $\rho_0 = 0.5$, and for $K = 12$.

Telecom Italia's customers in Milano, Italy, from 1/11/2013 to 1/1/2014. The spatial distribution λ_i of telecommunication events is aggregated in a 100×100 grid of locations $i \in \mathcal{I}$. The temporal distribution of events is aggregated over 10-minute time intervals. For our analysis we consider only working days, in order to evaluate the system under high traffic and under the periodicity created by the work-cycles behaviour of people. The used dataset consists of $P = 39$ days, each containing 144 time-slots, with a horizon of $T = 5616$ traffic observations.

Our network architecture consists of 40 BS, most of them being close to the city center, where the load is higher. We follow the setup of [7], and consider Macro- and Micro- BSS

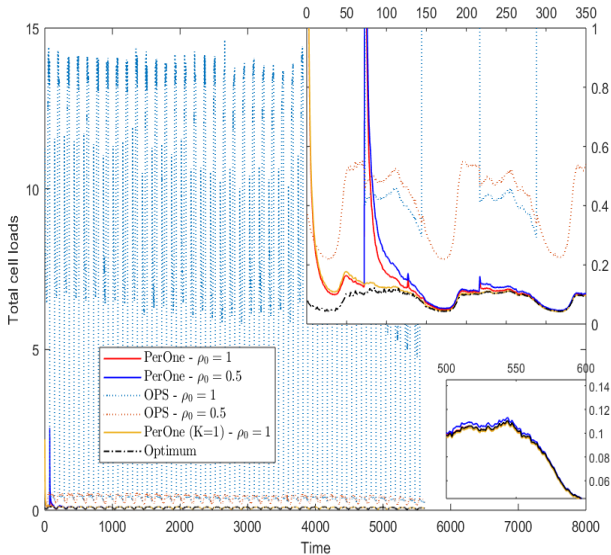


Fig. 5. Total cell loads vs. time, for $\alpha = 0$ and cost function $\phi_\alpha(\rho) = \sum_{j \in \mathcal{J}} \rho_j$ and for load threshold $\rho_0 = 1$ and $\rho_0 = 0.5$, and for $K = 2$.

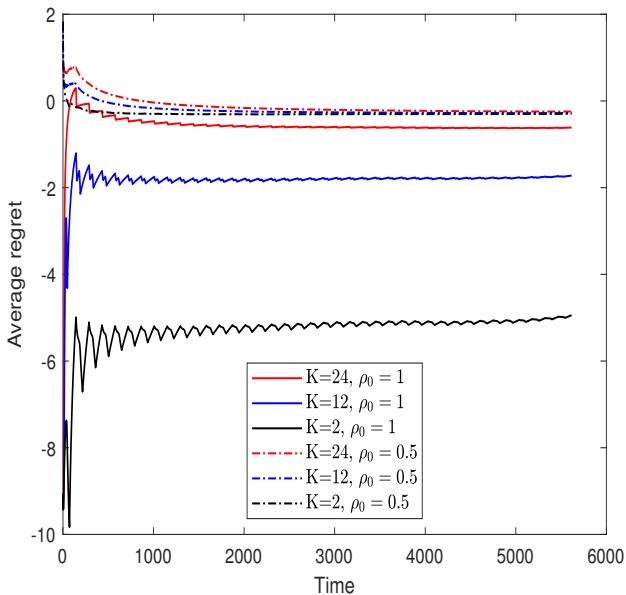


Fig. 6. PerOnE's regret over OPS, for $\alpha = 0$ and cost function $\phi_\alpha(\rho) = \sum_{j \in \mathcal{J}} \rho_j$ and for different values of ρ_0 and K .

transmitting at $P_M = 43\text{dBm}$ and $P_m = 33\text{dBm}$, respectively. The system bandwidth is $W = 10\text{MHz}$, while the noise density is $N_0 = -174\text{dBm/MHz}$. The path loss exponent is $P_{lo} = 3$, and G_{ji} is the resulting coefficient for the signal degradation from AP j to location i . Then, the transmission rates C_{ji} between location i and AP j are given by the Shannon formula:

$$C_{ji} = W \log_2 \left(1 + \frac{G_{ji} P_j}{W N_0 + \sum_{k \neq j} G_{ki} P_k} \right).$$

We are interested in evaluating the total cell loads that arise from user association policies produced by PerOnE, and to compare them to those of policies produced by OPS.

B. Results

We conduct a sensitivity analysis on the number K of time zones and load threshold ρ_0 values, to capture the scenario where the maximum available resources are considered ($\rho_0 = 1$), and a scenario with more limited resources ($\rho_0 = 0.5$). The "Optimum" is for OPS when $K = T$ and $\rho_0 = 1$, *i.e.*, it is the optimal association decision for each individual time slot t under the maximum amount of resources that could be considered. The PerOnE for $K = 1$ considers only one time zone, *i.e.*, runs taking as input the association policy of the previous slot, and without considering any division in the time horizon. For convenience, in our plots we provide an enlargement of the first time slots and of some slots that are indicative of how close PerOnE performs to the Optimum, at the top-right and bottom-right corner of the subfigures, respectively. We list some of our observations:

PerOnE quickly learns the optimal user association. PerOnE exploits the geometry of the problem and rapidly learns the optimal user association, despite the lack of actual traffic information. From Figs. 3-5, and Table II we see that as the number K of time zones increases, PerOnE needs more slots in order to learn not to violate constraints, to converge to optimal solutions and to produce more cost-efficient associations. This interesting feature allows PerOnE's solution updates to adapt to any traffic fluctuation. It is due to the fact that, as the number of the considered time zones decreases, the time slots that PerOnE initializes its decisions as for the expensive uniform solutions in (17) decreases too, similarly impacting the total cell loads. The contrary holds for OPS, whose static solutions benefit from a partition of the time horizon in more zones. Observe that it produces the minimum-cost static policies for a given partition of the time horizon, which does not necessarily imply that it will not have any constraint violations. In fact, as K grows, OPS violates constraints during more time slots and under more limited resources. However, observe from Figs. 3-5, that the actual cost of the produced associations grows as K decreases.

PerOnE effectively adapts to traffic changes. Despite the arbitrary and large traffic variations, PerOnE manages to adapt its solutions and decide cost-effective and near-optimal policies, both under high and low load threshold, as seen from Figs. 3-5. From these and table II, it can be observed that OPS fails to adapt, thus resulting to association policies that lead to a higher system load and constraint violations.

PerOnE has no regret against OPS. Despite the large fluctuations during the duration of the day, and the lack of actual information during the decision, PerOnE manages to produce asymptotically optimal solutions, under different partitions of each period in zones, and under different load-thresholds. As fig. 6 suggests, PerOnE's advantage over OPS grows as the load threshold ρ_0 increases and as the number K of time zones decreases. Intuitively, for larger ρ_0 has greater "margins" to adapt to the upcoming actual traffic. Also, for smaller K OPS is more restricted in its decisions, increasing PerOnE's advantage of adjusting its decisions dynamically.

VII. CONCLUSIONS

We assume arbitrary traffic variations over time. We introduce OPS, a novel periodic benchmark for online learning problems, which is significant to compare against in cases of conjectured traffic periodicity and generalizes state-of-the-art. We propose PerOnE, an asymptotically optimal online algorithm that produces association policies by performing a simple update. PerOnE learns to adapt to traffic fluctuations even under lack of actual information. We demonstrate its no-regret property against OPS both analytically and by performing simulations over a real-trace dataset. Moreover, PerOnE operates under no assumptions over traffic, which renders it a great user-association option for the highly dynamic environments envisioned for the large-scaled 5G and B5G/6G networks. In our future work we are interested to explore algorithms that jointly learn several dynamic parameters, for example user association and power control.

APPENDIX A PROOF OF LEMMA 1

The entropic function $h_i(\boldsymbol{\pi}) = \sum_{j \in \mathcal{J}} \pi_{ji} \log \pi_{ji}$ is 1-strongly convex with respect to the 1-norm [22], *i.e.*,

$$h_i(\boldsymbol{\pi}_1) \geq h_i(\boldsymbol{\pi}_2) + \langle \nabla h_i(\boldsymbol{\pi}_1), \boldsymbol{\pi}_2 - \boldsymbol{\pi}_1 \rangle + \frac{1}{2} \|\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1\|_1, \forall i \in \mathcal{I}.$$

Summing over all locations $i \in \mathcal{I}$ we obtain:

$$h(\boldsymbol{\pi}_1) \geq h(\boldsymbol{\pi}_2) + \sum_{i \in \mathcal{I}} \langle \nabla h_i(\boldsymbol{\pi}_1), \boldsymbol{\pi}_2 - \boldsymbol{\pi}_1 \rangle + \frac{|\mathcal{I}|}{2} \|\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1\|_1,$$

which due to the interchangeability of the sum and the dot-product, and due to (14), becomes:

$$h(\boldsymbol{\pi}_1) \geq h(\boldsymbol{\pi}_2) + \langle \nabla h(\boldsymbol{\pi}_1), \boldsymbol{\pi}_2 - \boldsymbol{\pi}_1 \rangle + \frac{|\mathcal{I}|}{2} \|\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1\|_1.$$

APPENDIX B PROOF OF THEOREM 1

For simplicity of notation, let $\mathbf{z}(t) := \nabla V(\boldsymbol{\pi}(t), \boldsymbol{\lambda}(t))$. Theorem 2.21 in [22] states that when the regularization function $h(\cdot)$ is $|\mathcal{I}|$ -strongly-convex w.r.t. a p -norm and the OMD is run with mirror function as in (13), then:

$$\sum_{t=1}^T \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^* \rangle \leq \frac{h(\boldsymbol{\pi}^*)}{\eta} - \frac{h(\boldsymbol{\pi}(1))}{\eta} + \sum_{t=1}^T \frac{\eta \|\mathbf{z}(t)\|_q^2}{2|\mathcal{I}|}, \quad (20)$$

where $\boldsymbol{\pi}^*$ an optimal static decision over T time slots as in (8), and the q -norm is the dual norm of the p -norm. We adopt this result and modify it to fit in the context of our periodic benchmark, using it for each time window \mathcal{W}_k , separately. Thus:

$$\sum_{t \in \mathcal{W}_k} \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^*[k] \rangle \leq \frac{h(\boldsymbol{\pi}^*[k])}{\eta} - \frac{h(\boldsymbol{\pi}(t_k^1))}{\eta} + \sum_{t \in \mathcal{W}_k} \frac{\eta \|\mathbf{z}(t)\|_q^2}{2|\mathcal{I}|}.$$

From the Lipschitz-continuity of the objective function, it exists a positive constant $L \geq \|\mathbf{z}(t)\|_q$, for all q -norms and time windows \mathcal{W}_k . Thus, for $k = 1, \dots, K$, it is:

$$\sum_{t \in \mathcal{W}_k} \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^*[k] \rangle \leq \frac{h(\boldsymbol{\pi}^*[k])}{\eta} - \frac{h(\boldsymbol{\pi}(t_k^1))}{\eta} + \sum_{t \in \mathcal{W}_k} \frac{\eta L^2}{2|\mathcal{I}|}, \quad (21)$$

Since the association variables for each location belong in $[0, 1]$, for each location i it holds that $h_i(\boldsymbol{\pi}) = \sum_{j \in \mathcal{J}} \pi_{ji} \log \pi_{ji} \leq 0$, which implies that

$$h(\boldsymbol{\pi}^*[k]) = \sum_{i \in \mathcal{I}} h_i(\boldsymbol{\pi}^*[k]) \leq 0. \quad (22)$$

Moreover:

$$\begin{aligned} h(\boldsymbol{\pi}(t_k^1)) &\stackrel{(14),(17)}{=} \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \frac{1}{|\mathcal{N}^i|} \log\left(\frac{1}{|\mathcal{N}^i|}\right) \\ &= - \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{N}_j} \frac{\log(|\mathcal{N}^i|)}{|\mathcal{N}^i|} \geq - \frac{M_I \log(M_J)}{M_J}, \forall k, \end{aligned} \quad (23)$$

where the inequality is due to (19), because $M_I \leq |\mathcal{I}|$ and $M_J \leq |\mathcal{J}|$. Then, (21) together with (22) and (23) leads to:

$$\sum_{t \in \mathcal{W}_k} \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^*[k] \rangle \leq \frac{M_I \log(M_J)}{\eta M_J} + |\mathcal{W}_k| \frac{\eta L^2}{2|\mathcal{I}|}, \forall k. \quad (24)$$

Convexity of the objective function implies:

$$V(\boldsymbol{\pi}(t), \boldsymbol{\lambda}(t)) - V(\boldsymbol{\pi}^*[k], \boldsymbol{\lambda}(t)) \leq \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^*[k] \rangle, \quad (25)$$

for all $t \in \mathcal{W}_k, k = 1, \dots, K$. Then:

$$\begin{aligned} \text{Reg}(T, K) &\stackrel{(7)}{=} \sum_{t=1}^T \phi_a(\boldsymbol{\pi}(t), \boldsymbol{\lambda}(t)) - \sum_{k=1}^K \sum_{t \in \mathcal{W}_k} \phi_a(\boldsymbol{\pi}^*[k], \boldsymbol{\lambda}(t)) \\ &\stackrel{(10),(6)}{\leq} \sum_{k=1}^K \sum_{t \in \mathcal{W}_k} (V(\boldsymbol{\pi}(t), \boldsymbol{\lambda}(t)) - V(\boldsymbol{\pi}^*[k], \boldsymbol{\lambda}(t))) \\ &\stackrel{(25)}{\leq} \sum_{k=1}^K \sum_{t \in \mathcal{W}_k} \langle \mathbf{z}(t), \boldsymbol{\pi}(t) - \boldsymbol{\pi}^*[k] \rangle \\ &\stackrel{(24)}{\leq} \sum_{k=1}^K \left(\frac{M_I \log(M_J)}{\eta M_J} + |\mathcal{W}_k| \frac{\eta L^2}{2|\mathcal{I}|} \right) \\ &= \frac{K M_I \log(M_J)}{\eta M_J} + \frac{\eta T L^2}{2|\mathcal{I}|}, \end{aligned} \quad (26)$$

which concludes the first part of the proof. The first equality is basically the definition of regret. From (6) $\boldsymbol{\pi}^*[k]$ minimizes costs for $t \in \mathcal{W}_k$, and from (10) the objective $V(\cdot)$ is at least equal to costs. The first inequality comes from the observation that the penalty paid by the optimal benchmark $\boldsymbol{\pi}^*[k]$ can't be greater than that paid by the online algorithm that takes decisions under lack of information. The second inequality comes as a result of the convexity of $V(\cdot)$, and the next from substituting the RHS of (24) for each window \mathcal{W}_k . The last equality holds because $\sum_{k=1}^K |\mathcal{W}_k| = T$, since summing all time slots over all the time windows is equivalent with summing over the entire time horizon.

For the second part of the Theorem, it is easily verifiable that (26) is minimized for $\eta = \sqrt{\frac{K M_I \log(M_J) 2|\mathcal{I}|}{T L^2 M_J}}$. Substituting in (26), and since $\log(M_J) < M_J$ and $M_I < |\mathcal{I}|$, we get:

$$\text{Reg}(T, K) \leq \sqrt{\frac{2 K M_I \log(M_J) T L^2}{M_J |\mathcal{I}|}} \leq L \sqrt{2 K T}.$$

REFERENCES

- [1] E. Calvanese Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6G: The Next Frontier: From Holographic Messaging to Artificial Intelligence Using Subterhertz and Visible Light Communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, 2019.

- [2] F. Kelly, A. Mauilloo, and D. Tan, "Rate control for communication networks: Shadow prizes, proportional fairness and stability," *Journal of Operation Research Society*, vol. 49, pp. 237–252, 1998.
- [3] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed α -optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. on Networking*, vol. 20, no. 1, pp. 177–190, Feb 2012.
- [4] L. Vigneri, G. Paschos, and P. Mertikopoulos, "Large-scale network utility maximization: Countering exponential growth with exponentiated gradients," in *IEEE Conference on Computer Communications - IEEE INFOCOM*, 2019.
- [5] B. Hajek, "Performance of global load balancing by local adjustment," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1398–1414, 1990.
- [6] M. Alanyali and B. Hajek, "On simple algorithms for dynamic load balancing," in *IEEE Conference on Computer Communications - IEEE INFOCOM*, vol. 1, 1995, pp. 230–238 vol.1.
- [7] N. Liakopoulos, GS. Paschos, and T. Spyropoulos, "Robust user association for ultra dense networks," in *IEEE Conference on Computer Communications - IEEE INFOCOM*, 2018, pp. 2690–2698.
- [8] I. Koutsopoulos and L. Tassiulas, "Joint optimal access point selection and channel assignment in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 521–532, 2007.
- [9] S. Papavassiliou and L. Tassiulas, "Improving the capacity in wireless networks through integrated channel base station and power assignment," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 417–427, 1998.
- [10] M. Karaliopoulos, L.E. Chatzieftheriou, G. Darzanos, and I. Koutsopoulos, "On the joint content caching and user association problem in small cell networks," in *2020 IEEE International Cnference on Communications - ICC Workshops*, 2020, pp. 1–6.
- [11] G. Darzanos, L.E. Chatzieftheriou, M. Karaliopoulos, and I. Koutsopoulos, "Content preference-aware user association and caching in cellular networks," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT) Workshops*, 2020, pp. 1–8.
- [12] L.E. Chatzieftheriou, G. Darzanos, M. Karaliopoulos, and I. Koutsopoulos, "Joint user association, content caching and recommendations in wireless edge networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 46, no. 3, pp. 12–17, 2018.
- [13] L. E. Chatzieftheriou, M. Karaliopoulos, and I. Koutsopoulos, "Caching-Aware Recommendations: Nudging User Preferences towards better Caching Performance," in *IEEE Conference on Computer Communications - IEEE INFOCOM*, 2017, pp. 784–792.
- [14] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *International Conference on Machine Learning - ICML*, 2003, pp. 928–935.
- [15] E. Hazan and S. Kale, "Projection-free online learning," in *International Conference on Machine Learning - ICML*, 2012.
- [16] H. Yu, M. Neely, and X. Wei, "Online convex optimization with stochastic constraints," in *Conference on Neural Information Processing Systems - NIPS*, 2017, pp. 1428–1438.
- [17] N. Liakopoulos, A. Destounis, G. Paschos, T. Spyropoulos, and P. Mertikopoulos, "Cautious regret minimization: Online optimization with long-term budget constraints," in *International Conference on Machine Learning - ICML*, 2019.
- [18] G. Paschos, A. Destounis, and G. Iosifidis, "Online convex optimization for caching networks," *IEEE/ACM Trans. on Networking*, 2020.
- [19] T. Karagkioules, GS. Paschos, N. Liakopoulos, A. Fiandrotti, D. Tsilimantos, and M. Cagnazzo, "Online learning for robust adaptive video streaming in mobile networks," *arXiv preprint: 1905.11705*, 2019.
- [20] T. Chen and Q. Ling and G. Giannakis, "An online convex optimization approach to proactive network resource allocation," *IEEE Trans. on Signal Processing*, 2017.
- [21] D. Bertsekas, *Convex Optimization Algorithms*. Athena Scientific, 2015.
- [22] S. Shalev-Shwartz, "Online Learning and Online Convex Optimization," *Foundations and Trends® in Machine Learning*, 2012.
- [23] E. V. Belmega, P. Mertikopoulos, R. Negrel, and L. Sanguinetti, "Online Convex Optimization and No-Regret Learning: Algorithms, Guarantees and Applications," *arXiv: 1804.04529*, 2018.
- [24] T. M. Cover, "Behavior of sequential predictors of binary sequences," in *Transactions on Prague Conference on Information Theory Statistical Decision Functions, Random Processes*, 1965, p. 263–272.
- [25] "Telecommunications - SMS, Call, Internet - ML," Telecom Italia, 2015. [Online]. Available: <https://doi.org/10.7910/DVN/EGZHFV>