Enabling Long-term Fairness in Dynamic Resource Allocation

TAREQ SI SALEM, Inria, France
GEORGE IOSIFIDIS, TU Delft, The Netherlands
GIOVANNI NEGLIA, Inria, France

We study the fairness of dynamic resource allocation problem under the α-fairness criterion. We recognize two different fairness objectives that naturally arise in this problem: the well-understood slot-fairness objective that aims to ensure fairness at every timeslot, and the less explored horizon-fairness objective that aims to ensure fairness across utilities accumulated over a time horizon. We argue that horizon-fairness comes at a lower price in terms of social welfare. We study horizon-fairness with the regret as a performance metric and show that vanishing regret cannot be achieved in presence of an unrestricted adversary. We propose restrictions on the adversary's capabilities corresponding to realistic scenarios and an online policy that indeed guarantees vanishing regret under these restrictions. We demonstrate the applicability of the proposed fairness framework to a representative resource management problem considering a virtualized caching system where different caches cooperate to serve content requests.

CCS Concepts: • **Applied computing → Multi-criterion optimization and decision-making**; • **Theory of computation → Online learning algorithms**; *Algorithmic game theory*.

Additional Key Words and Phrases: Online Learning, Multi-timescale Fairness, Axiomatic Bargaining, Dynamic Resource Allocation

ACM Reference Format:
Tareq Si Salem, George Iosifidis, and Giovanni Neglia. 2022. Enabling Long-term Fairness in Dynamic Resource Allocation. *Proc. ACM Meas. Anal. Comput. Syst.* 6, 3, Article 46 (December 2022), 36 pages. https://doi.org/10.1145/3570606

1 Introduction

Achieving fairness when allocating resources in communication and computing systems has been a subject of extensive research, and has been successfully applied in numerous practical problems. Fairness is leveraged to perform congestion control in the Internet [29, 53], to select transmission power in multi-user wireless networks [61, 77], and to allocate multidimensional resources in cloud computing platforms [19, 71, 74]. Depending on the problem at hand, the criterion of fairness can be expressed in terms of how the service performance is distributed across the end-users, or in terms of how the costs are balanced across the servicing nodes. The latter case exemplifies the natural link between fairness and load balancing in resource-constrained systems [43, 79]. A prevalent fairness metric is α-fairness, which encompasses the utilitarian principle (Bentham-Edgeworth solution [24]), proportional fairness (Nash bargaining solution [55]), max-min fairness (Kalai–Smorodinsky bargaining solution [41]), and, under some conditions, Walrasian equilibrium [31]. All these fairness metrics have been used in different cases for the design of resource management mechanisms [54, 58].

A common limitation of the above works is that they consider *static* environments. That is, the resources to be allocated and, importantly, the users' utility functions, are fixed and known to the

Authors' addresses: Tareq Si Salem, tareq.si-salem@inria.fr, Inria, Sophia Antipolis, France; George Iosifidis, G.Iosifidis@tudelft.nl, TU Delft, Delft, The Netherlands; Giovanni Neglia, giovanni.neglia@inria.fr, Inria, Sophia Antipolis, France.

decision maker. This assumption is very often unrealistic for today's communication and computing systems. For instance, in small-cell mobile networks the user churn is typically very high and unpredictable, thus hindering the fair allocation of spectrum to cells [37]. Similarly, placing content files at edge caches to balance the latency gains across the served areas is non-trivial due to the non-stationary and fast-changing patterns of requests [30]. At the same time, the increasing virtualization of these systems introduces cost and performance volatility, as extensive measurement studies have revealed [26, 44, 73]. This uncertainty is exacerbated for services that process user-generated data (e.g., streaming data applications) where the performance (e.g., inference accuracy) depends also on a priori unknown input data and dynamically selected machine learning libraries [2, 40, 47].

1.1 Contributions

This paper makes the next step towards enabling long-term fairness in dynamic systems. We consider a system that serves a set of agents $\mathcal{I}$, where a controller selects at each timeslot $t \in \mathbb{N}$ a resource allocation profile $\boldsymbol{x}_t$ from a set of eligible allocations $\mathcal{X}$ based on past agents' utility functions $\boldsymbol{u}_{t'} : \mathcal{X} \to \mathbb{R}^{\mathcal{I}}$ for $t' < t$ and of α-fairness function $F_\alpha : \mathbb{R}^{\mathcal{I}}_{\geq 0} \to \mathbb{R}$. The utilities might change due to unknown, unpredictable, and (possibly) non-stationary perturbations that are revealed to the controller only after it decides $\boldsymbol{x}_t$. We employ the terms *horizon-fairness* (HF) and *slot-fairness* (SF) to distinguish the different ways fairness can be enforced in a such time-slotted dynamic system. Under horizon-fairness, the controller enforces fairness on the aggregate utilities for a given time horizon T, whereas under slot-fairness, it enforces fairness on the utilities at each timeslot separately. Both metrics have been studied in previous work, e.g., see [32, 38, 46, 69] and the discussion in Section 2. Our focus is on horizon-fairness, which raises novel technical challenges and subsumes slot-fairness as a special case.

We design the *online horizon-fair* (OHF) policy by leveraging *online convex optimization* (OCO) [34], to handle this reduced-information setting under a powerful *adversarial* perturbation model. Adversarial analysis is a modeling technique to characterize a system's performance under unknown and hard to characterize exogenous parameters and has been recently successfully used to model caching problems (e.g., in [18, 45, 52, 56, 57, 66, 67]). In our context, the performance of a resource allocation policy $\mathcal{A}$ is evaluated by the *fairness regret*, which is defined as the difference between the α-fairness, over the time-averaged utilities, achieved by a static optimum-in-hindsight (*benchmark*) and the one achieved by the policy:

$$\mathfrak{R}_T\left(F_\alpha, \mathcal{A}\right) \triangleq \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ \max_{\boldsymbol{x} \in \mathcal{X}} F_\alpha\left(\frac{1}{T}\sum_{t \in \mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x})\right) - F_\alpha\left(\frac{1}{T}\sum_{t \in \mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_t)\right)\right\}. \tag{1}$$

If the fairness regret vanishes over time (i.e., $\lim_{T\to\infty} \mathfrak{R}_T\left(F_\alpha, \mathcal{A}\right) = 0$), policy $\mathcal{A}$ will attain the same fairness value as the static benchmark under any possible sequence of utility functions. A policy that achieves sublinear regret under these adversarial conditions, can also succeed in more benign conditions where the perturbations are not adversarial, or the utility functions are revealed at the beginning of each slot.

The fairness regret metric (1) departs from the template of OCO. In particular, the scalarization of the vector-valued utilities, through the α-fairness function, is not applied at every timeslot to allow for the controller to easily adapt its allocations, instead is only applied at the end of the time horizon T. Our first result characterizes the challenges in tackling this learning problem. Namely, Theorem 1 proves that, when utility perturbations are only subject to four mild technical conditions, such as in standard OCO, it is impossible to achieve vanishing fairness-regret. Similar negative results were obtained under different setups of primal-dual learning and online saddle point learning [5, 50, 62], but they have been devised for specific problem structures (e.g., online matrix games) and thus do not apply to our setting.

In light of this negative result, we introduce additional *necessary* conditions on the adversary to obtain a vanishing regret guarantee. Namely, the adversary can only induce perturbations to the time-averaged utilities we call budgeted-severity or partitioned-severity constrained. These conditions capture several practical utility patterns, such as non-stationary corruptions, ergodic and periodic inputs [9, 23, 46, 80]. We proceed to propose the OHF policy which adapts dynamically the allocation decisions and provably achieves $\mathfrak{R}_T(F_\alpha, \boldsymbol{\mathcal{A}}) = o(1)$ (see Theorem 2).

The OHF policy employs a novel learning approach that operates concurrently, and in a synchronized fashion, in a primal and a dual (conjugate) space. Intuitively, OHF learns the weighted time-varying utilities in a primal space, and learns the weights accounting for the global fairness metric in some dual space. To achieve this, we develop novel techniques through a convex conjugate approach (see Lemmas 1, 2, and 4 in the Appendix).

Finally, we apply our fairness framework to a representative resource management problem in virtualized caching systems where different caches cooperate by serving jointly the received content requests. We evaluate the performance of OHF with its slot-fairness counterpart policy through numerical examples. We evaluate the price of fairness of OHF, which quantifies the efficiency loss due to fairness, across different network topologies and participating agents. Lastly, we apply OHF to a Nash bargaining scenario, a concept that has been widely used in resource allocation to distribute to a set of agents the utility of their cooperation [17, 36, 39, 75].

1.2 Outline of Paper

The paper is organized as follows. The related literature is discussed in Section 2. The definitions and background are provided in Section 3. The adversarial model and the proposed algorithm are presented in Section 4. Extensions to the fairness framework are provided in Section 5. The resource management problem in virtualized caching systems application is provided in Section 6. Finally, we conclude the paper and provide directions for future work in Section 7.

2 Literature Review

2.1 Fairness in Resource Allocation

Fairness has found many applications in wired and wireless networking [3, 29, 53, 61, 77], and cloud computing platforms [19, 71, 74]. Prevalent fairness criteria are the max-min fairness and proportional fairness, which are rooted in axiomatic bargaining theory, namely the Kalai–Smorodinsky [41] and Nash bargaining solution [55], respectively. On the other side of the spectrum, a controller might opt to ignore fairness and maximize the aggregate utility of users, i.e., to follow the *utilitarian principle*, also referred to as the Bentham-Edgeworth solution [24]. The *Price of Fairness* (PoF) [13] is now an established metric for assessing how much the social welfare (i.e., the aggregate utility) is affected when enforcing some fairness metric. Ideally, we would like this price to be as small as possible, bridging in a way these two criteria. Atkinson [6] proposed the unifying α-fairness criterion which yields different fairness criteria based on the value of $\alpha \in \mathbb{R}_{\geq 0}$, i.e., the utilitarian principle ($\alpha = 0$), proportional fairness ($\alpha = 1$), and max-min fairness ($\alpha \to \infty$). Due to the generality of the α-fairness criterion, we use it to develop our theory, which in turn renders our results transferrable to all above fairness and bargaining problems. In this work, the PoF, together with the metric of fairness-regret, are the two criteria we use to characterize our fairness solution.

2.2 Fairness in Dynamic Resource Allocation

Several works consider slot-fairness in dynamic systems [38, 69, 72]. Jalota and Ye [38] proposed a weighted proportional fairness algorithm for a system where new users arrive in each slot, having linear i.i.d. perturbed unknown utility functions at the time of selecting an allocation, and are allocated resources from an i.i.d. varying budget. Sinclair et al. [69] consider a similar setup, but assume the utilities are known at the time of selecting an allocation, and the utility parameters (number of agents and their type) are drawn from some fixed known distribution. They propose an adaptive threshold policy, which achieves a target efficiency (amount of consumed resources' budget) and fairness tradeoff, where the latter is defined w.r.t. to an offline weighted proportional fairness benchmark. Finally, Talebi and Proutiere [72] study dynamically arriving tasks that are assigned to a set of servers with unknown and stochastically-varying service rates. Using a stochastic multi-armed bandit model, the authors achieve proportional fairness across the service rates assigned to different tasks at each slot. All these important works, however, do not consider the more practical horizon-fairness metric where fairness is enforced throughout the entire operation of the system and not over each slot separately.

Horizon-fairness has been recently studied through the lens of competitive analysis [10, 11, 42], where the goal is to design a policy that achieves online fairness within a constant factor from the fairness of a suitable benchmark. Kawase and Sumita [42] consider the problem of allocating arriving items irrevocably to one agent who has additive utilities over the items. The arrival of the items is arbitrary and can even be selected by an adversary. The authors consider known utility at the time of allocation, and design policies under the max-min fairness criterion. Banerjee et al. [10] consider a similar problem under the proportional fairness criterion, and they allow the policies to exploit available predictions. We observe that the competitive ratio guarantees, while theoretically interesting, may not be informative about the fairness of the actual approximate solution achieved by the algorithm for ratios different from one. For instance, when maximizing a Nash welfare function under the proportional fairness criterion, the solution achieves some axiomatic fairness properties [55] (e.g., Pareto efficiency, individual rationality, etc.), but this welfare function is meaningless for "non-optimal" allocations [69], i.e., a policy with a high competitive ratio is not necessary less fair than a policy with a lower competitive ratio. For this reason, our work considers regret as a performance metric: when regret vanishes asymptotically, the allocations of the policy indeed achieve the exact same objective as the adopted benchmark.

Altman et al. [4] consider the α-fairness problem in a dynamic resource allocation, and investigate fairness enforced at different time scales (instantaneous and long-term). They consider known utilities at the time of selecting an allocation in a stationary setting. Lodi et al. [49] also treat fairness across different time scales (single-period and T-period) as an offline problem. In this work, we make a similar distinction on the fairness criterion in the online setting, where we define the slot-fairness and horizon-fairness.

A different line of work [8, 12, 20, 32, 46, 70, 78] considers horizon-fairness through regret analysis. Gupta and Kamble [32] study individual fairness criteria that advocate similar individuals should be treated similarly. They extend the notion of individual fairness to online contextual decision-making, and introduce: (1) fairness-across-time and (2) fairness-in-hindsight. Fairness-across-time criterion requires the treatment of individuals to be individually fair relative to the past as well as future, while fairness-in-hindsight only requires individual fairness at the time of the decision. The utilities are known at the time of selecting an allocation and are i.i.d. and drawn from an unknown fixed distribution. Liao et al. [46] consider a similar setup to ours, with a limited adversarial model and time-varying but known utilities, and focus on proportional fairness. They consider adversarial perturbation added on a fixed item distribution where the demand of

Table 1. Summary of related work under online fairness in resource allocation.

Paper	Criterion	HF/SF	Unknown utilities	Adversarial utilities	Metric
[38]	Weighted proportional fairness	SF	✓	×	Regret
[69]	Weighted proportional fairness	SF	×	×	Envy, Efficiency
[72]	Proportional fairness	SF	×	×	Regret
[32]	Individual fairness	HF/SF	×	×	Regret
[46]	Proportional fairness	HF	×	✓	Regret
[20]	α-fairness	HF	✓	×	Regret
[12]	Envy-freeness	HF	×	✓	Envy
[78]	Weighted proportional fairness	HF	×	✓	Envy, Pareto Efficiency
[8]	Proportional fairness	HF	×	×	Regret
[42]	Max-Min fairness	HF	×	✓	Competitive ratio
[10]	Proportional fairness	HF	×	✓	Competitive ratio
[11]	Proportional fairness	HF	×	×	Competitive ratio
This work	Weighted α-fairness	HF/SF	✓	✓	Fairness Regret

items generally behaves predictably, but for some time steps, the demand behaves erratically. Our approach departs significantly from these interesting works in that we consider unknown utility functions, a broader adversarial model (in fact, as broad as possible while still achieving vanishing fairness regret), and by using the general α-fairness criterion that encompasses all the above criteria as special cases. This makes, we believe, our OHF algorithm applicable to a wider range of practical problems. Table 1 summarizes the differences between our contribution and the related works.

2.3 Online Learning

Achieving horizon-fairness in our setup requires technical extensions to the theory of OCO [34]. The basic template of OCO-learning (in terms of resource allocation) considers that a decision maker selects repeatedly a vector $\boldsymbol{x}_t$ from a convex set $\mathcal{X}$, before having access to the t-th slot scalar utility function $u_t(\boldsymbol{x})$, with the goal to maximize the aggregate utility $\sum_{t=1}^{T} u_t(\boldsymbol{x}_t)$. The decision maker aims to have vanishing time-averaged regret, i.e., the time-averaged distance of the aggregate utility $\sum_{t=1}^{T} u_t(\boldsymbol{x}_t)$ from the aggregate utility of the optimal-in-hindsight allocation $\max_{\boldsymbol{x}\in\mathcal{X}} \sum_{t=1}^{T} u_t(\boldsymbol{x})$ for some time horizon T. OCO models are robust, expressive, and can be tackled with several well-studied learning algorithms [34, 51, 63]. However, none of those is suitable for the fairness problem at hand, as we need to optimize a global function $F_\alpha(\cdot)$ of the time-averaged vector-valued utilities. This subtle change creates additional technical complications. Indeed, optimizing functions of time-averaged utility/cost functions in learning is an open and challenging problem. In particular, Even-Dar et al. [28] introduce the concept of global functions in online learning, and devise a policy with vanishing regret using the theory of approachability [15]. However, their approach can handle only norms as global functions, and this limitation is not easy to overcome: the authors themselves stress that characterizing when a global function enables a vanishing regret is an open problem (see [28, Section 7]). Rakhlin et al. [60] extend this work to non-additive global functions. However, the α-fairness function considered in our work is not supported by their framework. To generalize the results to α-fairness global functions, we employ a convex conjugate approach conceptually similar to the approach taken in the work of Agrawal and Evanur [1] to obtain a regret guarantee with a concave global function under a stationary setting and linear utilities. In this work, we consider an adversarial setting (i.e., utilities are picked by an adversary after we select an allocation) that encompasses general concave utilities, and this requires learning over the primal space as well as the dual (conjugate) space.

3 Online Fairness: Definitions and Background

3.1 Static Fairness

Consider a system $\mathcal{S}$ that serves a set of agents $\mathcal{I}$ by selecting allocations from the set of eligible allocations $\mathcal{X}$.[1] In the general case, this set is defined as the Cartesian product of agent-specific eligible allocations' set $\mathcal{X}_i$, i.e., $\mathcal{X} \triangleq \bigtimes_{i\in\mathcal{I}} \mathcal{X}_i$. We assume that each set $\mathcal{X}_i$ is convex. The utility of each agent $i \in \mathcal{I}$ is a concave function $u_i : \mathcal{X} \to \mathbb{R}_{\geq 0}$, and depends, possibly, not only on $\boldsymbol{x}_i \in \mathcal{X}_i$, but on the entire vector $\boldsymbol{x} \in \mathcal{X}$.[2] The vector $\boldsymbol{u}(\boldsymbol{x}) \triangleq (u_i(\boldsymbol{x}))_{i\in\mathcal{I}} \in \mathcal{U}$ is the vectorized form of the agents' utilities, where $\mathcal{U}$ is the set of possible utility functions. The joint allocation $\boldsymbol{x}_\star \in \mathcal{X}$ is an α-fair allocation for some $\alpha \in \mathbb{R}_{\geq 0}$ if it solves the following convex problem:

$$\max_{\boldsymbol{x}\in\mathcal{X}} F_\alpha\left(\boldsymbol{u}(\boldsymbol{x})\right), \tag{2}$$

where F_α is the α-fairness criterion the system employs (e.g., when $\alpha = 1$, problem (2) corresponds to an Eisenberg-Gale convex problem [25]). The α-fairness function is defined as follows [6]:

Definition 1. *An α-fairness function $F_\alpha : \mathcal{U} \to \mathbb{R}$ is parameterized by the inequality aversion parameter $\alpha \in \mathbb{R}_{\geq 0}$, and it is given by*

$$F_\alpha(\boldsymbol{u}) \triangleq \sum_{i\in\mathcal{I}} f_\alpha(u_i), \qquad \text{where} \qquad f_\alpha(u) \triangleq \begin{cases} \frac{u^{1-\alpha}-1}{1-\alpha}, & \text{for } \alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}, \\ \log(u), & \text{for } \alpha = 1, \end{cases} \tag{3}$$

for every $\boldsymbol{u} \in \mathcal{U}$. Note that $\mathcal{U} \subset \mathbb{R}^{\mathcal{I}}_{\geq 0}$ for $\alpha < 1$, and $\mathcal{U} \subset \mathbb{R}^{\mathcal{I}}_{>0}$ for $\alpha \geq 1$.

Note that we use the most general version of utility-based fairness where the fairness rule is defined w.r.t. to accrued utilities (as opposed to allocated resource, only), i.e., in our system $\mathcal{S}$, the utility vector $\boldsymbol{u} \in \mathcal{U}$ can be a function of the selected allocations in $\mathcal{X}$. The α-fairness function is concave and component-wise increasing, and thus exhibits diminishing returns [14]. An increase in utility to a player with a low utility results in a higher α-fairness objective. Thus, such an increase is desirable to the system controller. Moreover, the rate at which the marginal increase diminishes is controlled by α, which is then called the *inequality aversion parameter*. An allocation which maximizes the α-fairness objective is always Pareto efficient [14].

3.2 Online Fairness

We consider the performance of the system $\mathcal{S}$ is tracked over a time horizon spanning $T \in \mathbb{N}$ timeslots. At the beginning of each timeslot $t \in \mathcal{T} \triangleq \{1, 2, \ldots, T\}$, a policy selects an allocation $\boldsymbol{x}_t \in \mathcal{X}$ *before* $\boldsymbol{u}_t : \mathcal{X} \to \mathbb{R}^{\mathcal{I}}$ is revealed to the policy. The goal is to approach the performance of a properly-selected fair allocation benchmark. We consider the following two cases:

Slot-Fairness. An offline benchmark in hindsight, with access to the utilities revealed at every timeslot $t \in \mathcal{T}$, can ensure fairness at every timeslot satisfying a *slot-fairness* (SF) objective [38, 69, 72]. Formally, the benchmark selects the joint allocation $\boldsymbol{x}_\star \in \mathcal{X}$ satisfying

$$\text{SF}: \qquad \boldsymbol{x}_\star \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{T} \sum_{t\in\mathcal{T}} F_\alpha\left(\boldsymbol{u}_t(\boldsymbol{x})\right). \tag{4}$$

[1] Appendix G discusses the setting in which the set of agents $\mathcal{I}$ is unknown and agents can depart and arrive to the system.
[2] For example, in TCP congestion control, the performance of each end-node depends not only on the rate that is directly allocated to that node, but also, through the induced congestion in shared links, by the rate allocated to other nodes [29]. Similar couplings arise in wireless transmissions over shared channels [61].

Horizon-Fairness. Enforcing fairness at every timeslot can be quite restrictive, and this is especially evident for large time horizons. An alternative formulation is to consider that the agents can accept a momentary violation of fairness at a given timeslot $t \in \mathcal{T}$ as long as in the long run fairness over the total incurred utilities is achieved. Therefore, it is more natural (see Example 1) to ensure a horizon-fairness criterion over the entire period $\mathcal{T}$. Formally, the benchmark selects the allocation $\boldsymbol{x}_\star \in \mathcal{X}$ satisfying

$$\text{HF}: \qquad \boldsymbol{x}_\star \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} F_\alpha\left(\frac{1}{T}\sum_{t\in\mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x})\right). \tag{5}$$

Price of fairness. Bertsimas et al. [14] defined the *price of fairness* (PoF) metric to quantify the efficiency loss due to fairness as the difference between the maximum system efficiency and the efficiency under the fair scheme. In the case of α-fairness, it is defined for some utility set $\mathcal{U}$ as

$$\text{PoF}(\mathcal{U};\alpha) \triangleq \frac{\max_{\boldsymbol{u}\in\mathcal{U}} F_0(\boldsymbol{u}) - F_0\left(\boldsymbol{u}_{\max,\alpha}\right)}{\max_{\boldsymbol{u}\in\mathcal{U}} F_0(\boldsymbol{u})}, \tag{6}$$

where $\boldsymbol{u}_{\max,\alpha} \in \arg\max_{\boldsymbol{u}\in\mathcal{U}} F_\alpha(\boldsymbol{u})$ and $F_0(\boldsymbol{u}) = \sum_{i\in\mathcal{I}} u_i$ measures the achieved social welfare. Note that by definition the utilitarian objective achieves maximum efficiency, i.e., $\text{PoF}(\mathcal{U};0) = 0$. Naturally, in our online setting, the metric is extended as follows

$$\text{PoF}(\mathcal{X};\mathcal{T};\alpha) \triangleq \frac{\max_{\boldsymbol{x}\in\mathcal{X}} \sum_{t\in\mathcal{T}} F_0\left(\boldsymbol{u}_t(\boldsymbol{x})\right) - \sum_{t\in\mathcal{T}} F_0\left(\boldsymbol{u}_t(\boldsymbol{x}_\star)\right)}{\max_{\boldsymbol{x}\in\mathcal{X}} \sum_{t\in\mathcal{T}} F_0\left(\boldsymbol{u}_t(\boldsymbol{x})\right)}, \tag{7}$$

where $\boldsymbol{x}_\star$ is obtained through either SF (4) or HF (5). We provide the following example to further motivate our choice of horizon-fairness as a performance objective. A similar argument is provided in [49, Example 7].

Example 1. Consider a system with two agents $\mathcal{I} = \{1, 2\}$, an allocation set $\mathcal{X} = [0, x_{\max}]$ with $x_{\max} > 1$, α-fairness criterion with $\alpha = 1$, even $T \in \mathbb{N}$, and the following sequence of utilities $\{\boldsymbol{u}_t(x)\}_{t=1}^{T} = \{(1+x, 1-x), (1+x, 1+x), \dots\}$. It can easily be verified that PoF $= 0$ for HF objective (5) because the HF optimal allocation is $x_{\max}$ which matches the optimal allocation under the utilitarian objective. However, under the SF objective (4) we have PoF $= \frac{x_{\max}-0.5}{x_{\max}+2} \approx 1$ when $x_{\max}$ is large. Remark that the two objectives have different domains of definitions; in particular, the allocations in the set $[1, x_{\max}] \subset \mathcal{X}$ are unachievable by the SF objective because they would lead to $u_{t,2}(x) \leq 0$. The HF objective achieves lower PoF (hence, larger aggregate utility), and it allows a much larger set of eligible allocations (in particular all the allocations in the set $\mathcal{X}$), as shown in Fig. 1. Indeed, when the controller has the freedom to achieve fairness over a time horizon, there is an opportunity for more efficient allocations during the system operation. This example provides intuition on the robustness and practical importance of the horizon-fairness objective.

In the following section, we provide the description of an online learning model and our performance metric of interest under the HF objective.

3.3 Online Policies and Performance Metric

The agents' allocations are determined by an online policy $\boldsymbol{\mathcal{A}} = \{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_T\}$, i.e., a sequence of mappings. For every timeslot $t \in \mathcal{T}$, $\mathcal{A}_t : \mathcal{X}^t \times \mathcal{U}^t \to \mathcal{X}$ maps the sequence of past allocations $\{\boldsymbol{x}_s\}_{s=1}^{t} \in \mathcal{X}^t$ and utility functions $\{\boldsymbol{u}_s\}_{s=1}^{t} \in \mathcal{U}^t$ to the next allocation $\boldsymbol{x}_{t+1} \in \mathcal{X}$. We assume the initial decision $\boldsymbol{x}_1$ is feasible (i.e., $\boldsymbol{x}_1 \in \mathcal{X}$). We measure the performance of policy $\boldsymbol{\mathcal{A}}$ in terms of the *fairness regret* (8), i.e., the difference between the fairness objective experienced by $\boldsymbol{\mathcal{A}}$ at the time horizon T and that of the best static decision $\boldsymbol{x}_\star \in \mathcal{X}$ in hindsight. We restate the regret metric

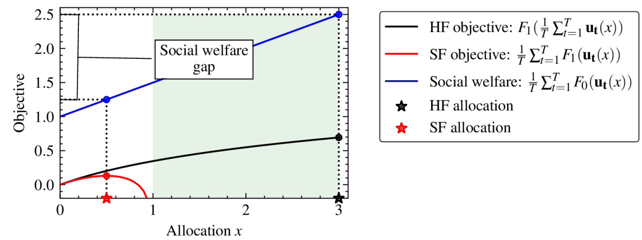


Fig. 1. Price of Fairness under HF and SF objectives for Example 1 for $x_{\max} = 3$. The green shaded area provides the set of allocation unachievable by the SF objective but achievable by the HF objective.

here to streamline the presentation:

$$\mathfrak{R}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) \triangleq \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ F_\alpha\left(\frac{1}{T}\sum_{t\in\mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_\star)\right) - F_\alpha\left(\frac{1}{T}\sum_{t\in\mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_t)\right)\right\}. \tag{8}$$

where $\boldsymbol{x}_\star$ is the HF (5) allocation. If the fairness regret becomes negligible for large T, then $\boldsymbol{\mathcal{A}}$ attains the same fairness objective as the optimal static decision with hindsight. Note that under the utilitarian objective ($\alpha = 0$), this fairness regret coincides with the classic time-averaged regret in OCO [34]. However, for general values of $\alpha \neq 0$, the metric is completely different, as we aim to compare α-fair functions evaluated at time-averaged vector-valued utilities.

4 Online Horizon-Fair (OHF) Policy

We first present in Section 4.1, the adversarial model considered in this work and provide a result on the impossibility of guaranteeing vanishing fairness regret (8) under general adversarial perturbations. We also provide a powerful family of adversarial perturbations for which a vanishing fairness regret guarantee is attainable. Secondly, we present the OHF policy in Section 4.2 and provide its performance guarantee. Finally, we provide in Section 4.3 a set of adversarial examples captured by our fairness framework.

4.1 Adversarial Model and Impossibility Result

We begin by introducing formally the adversarial model that characterizes the utility perturbations. In particular, we consider $\boldsymbol{\delta}_t(\boldsymbol{x}) \triangleq \left(\frac{1}{T}\sum_{s\in\mathcal{T}}\boldsymbol{u}_s(\boldsymbol{x})\right) - \boldsymbol{u}_t(\boldsymbol{x})$ to quantify how much the adversary *perturbs* the average utility by selecting a utility function $\boldsymbol{u}_t$ at timeslot $t \in \mathcal{T}$. Recall that $\boldsymbol{x}_\star \in \mathcal{X}$ denotes the optimal allocation under HF objective (5). We denote by $\Xi(\mathcal{T})$ the set of all possible decompositions of $\mathcal{T}$ into sets of contiguous timeslots, i.e., for every $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_K\} \in \Xi(\mathcal{T})$ it holds $\mathcal{T} = \dot{\bigcup}_{k\in\{1,2,\dots,K\}}\mathcal{T}_k$ and $\max\mathcal{T}_k < \min\mathcal{T}_{k+1}$ for $k \in \{1, 2, \dots, K-1\}$. We define two types of adversarial perturbations:

$$\text{Budgeted-severity:} \quad \mathbb{V}_{\mathcal{T}} \triangleq \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}} \left|\delta_{t,i}(\boldsymbol{x}_\star)\right| \right\}, \tag{9}$$

$$\text{Partitioned-severity:} \quad \mathbb{W}_{\mathcal{T}} \triangleq \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ \inf_{\substack{\{\mathcal{T}_1,\mathcal{T}_2,\dots,\mathcal{T}_K\}\\ \in\Xi(\mathcal{T})}} \left\{ \sum_{k=1}^{K}\sum_{i\in\mathcal{I}} \left|\sum_{t\in\mathcal{T}_k}\delta_{t,i}(\boldsymbol{x}_\star)\right| + \sum_{k=1}^{K} \frac{|\mathcal{T}_k|^2}{\sum_{k'<k}|\mathcal{T}_k| + 1} \right\}\right\}. \tag{10}$$

Our result in Theorem 2 implies that when either $\mathbb{V}_{\mathcal{T}}$ or $\mathbb{W}_{\mathcal{T}}$ grows sublinearly in the time horizon (i.e., the perturbations satisfy at least one of these two conditions), the regret of OHF policy in Algorithm 1 vanishes over time. We provide a detailed description of conditions (9) and (10) below.

The *budgeted-severity* $\mathbb{V}_{\mathcal{T}}$ in Eq. (9) bounds the total amount of perturbations of the time-averaged utility. When $\mathbb{V}_{\mathcal{T}} = 0$ the adversary is only able to select a fixed function, otherwise the adversary is able to select time-varying utilities, while keeping the total deviation no more than $\mathbb{V}_{\mathcal{T}}$. Moreover, the adversary is allowed to pick opportunely the timeslots to maximize performance degradation for the controller. This model is similar to the *adversarial corruption* setting considered in [9, 46], and it captures realistic scenarios where the utilities incurred at different timeslots are predictable, but can be perturbed for some fraction of the timeslots. For instance, Internet traffic may experience spikes due to breaking news or other unpredictable events [27].

The *partitioned-severity* $\mathbb{W}_{\mathcal{T}}$ in Eq. (10) may at first be less easy to understand than budgeted-severity condition (9), but is equally important from a practical point of view. For simplicity, consider a uniform decomposition of the timeslots, i.e., $\mathcal{T}_k = M$ for every $k \in \{1, 2, \dots, T/M\}$ assuming w.l.g. M divides T. Then the r.h.s. term in Eq. (10) can be bounded as follows:

$$\sum_{k=1}^{T/M} \frac{|\mathcal{T}_k|^2}{\sum_{k'<k} |\mathcal{T}_k| + 1} = \sum_{k=1}^{T/M} \frac{M^2}{M(k-1)+1} = \mathcal{O}\left(M^2 + M\log(T/M)\right). \tag{11}$$

Hence, when $M = o(\sqrt{T})$ it holds $\sum_{k=1}^{T/M} \frac{|\mathcal{T}_k|^2}{\sum_{k'<k}|\mathcal{T}_k|+1} = o(T)$. Since this term grows sublinearly in time, it remains to characterize the growth of the l.h.s. term $\sum_{k=1}^{K}\sum_{i\in\mathcal{I}}\left|\sum_{t\in\mathcal{T}_k}\delta_{t,i}(\boldsymbol{x}_\star)\right|$ in Eq. (10). This term is related to the perturbations selected by the adversary, however the absolute value is only evaluated at the end of each contiguous subperiod $\mathcal{T}_k$, i.e., the positive and negative deviations from the average utilities can cancel out. For example, a periodic selection of utilities from some set with cardinality M would have zero deviation for this term. This type of adversary is similar to the periodic adversary considered in [9, 23], but also includes adversarial selection of utilities from some finite set (see Example 3 in Section 4.3). The partitioned-severity adversary can model real-life applications that exhibit seasonal properties, e.g., the traffic may be completely different throughout the day, but daily traffic is self-similar [80]. This condition also unlocks the possibility to obtain high probability guarantees under stochastic utilities (see Corollary 4).

We formally make the following assumptions:

- (A1) The allocation set $\mathcal{X}$ is convex with diameter $\operatorname{diam}(\mathcal{X}) < \infty$.
- (A2) The utilities are bounded, i.e., $\boldsymbol{u}_t(\boldsymbol{x}) \in [u_{\min}, u_{\max}]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}$ for every $t \in \mathcal{T}$.
- (A3) The supergradients of the utilities are bounded over $\mathcal{X}$, i.e., it holds $\|\boldsymbol{g}\|_2 \leq L_{\mathcal{X}} < \infty$ for any $\boldsymbol{g} \in \partial_{\boldsymbol{x}} u_{t,i}(\boldsymbol{x})$ and $\boldsymbol{x} \in \mathcal{X}$.
- (A4) The average utility of the optimal allocation (5) is bounded such that $\frac{1}{T}\sum_{t\in\mathcal{T}}\boldsymbol{u}_t(\boldsymbol{x}_\star) \in \left[u_{\star,\min}, u_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{>0}$.
- (A5) The adversary is restricted to select utilities such that

$$\min\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\} = o(T). \tag{12}$$

We first show that an adversary solely satisfying the mild assumptions (A1)–(A4) can arbitrarily degrade the performance of any policy $\boldsymbol{\mathcal{A}}$. Formally, we have the following negative result:

Theorem 1. *When Assumptions (A1)–(A4) are satisfied, there is no online policy $\boldsymbol{\mathcal{A}}$ attaining $\mathfrak{R}_T(F_\alpha, \boldsymbol{\mathcal{A}}) = \mathrm{o}(1)$ for $|\mathcal{I}| > 1$ and $\alpha > 0$. Moreover, there exists an adversary where Assumption (A5) is necessary for $\mathfrak{R}_T(F_\alpha, \boldsymbol{\mathcal{A}}) = o(1)$.*

The proof can be found in Appendix B. We design an adversary with a choice over two sequences of utilities against two agents. We show that no policy can have vanishing fairness regret w.r.t. the time horizon under both sequences.

Algorithm 1 OHF policy

Require: $\mathcal{X}, \alpha \in \mathbb{R}_{\geq 0}, \left[u_{\star,\min}, u_{\star,\max}\right]$

1: $\Theta \leftarrow \left[-1/u^{\alpha}_{\star,\min}, -1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}}$ ▷ Initialize the dual (conjugate) subspace

2: $\boldsymbol{x}_1 \in \mathcal{X}; \boldsymbol{\theta}_1 \in \Theta;$ ▷ Initialize allocation $\boldsymbol{x}_1$ and dual decision $\boldsymbol{\theta}_1$

3: **for** $t \in \mathcal{T}$ **do**

4: Reveal $\Psi_{t,\alpha}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) = (-F_\alpha)^\star(\boldsymbol{\theta}_t) - \boldsymbol{\theta}_t \cdot \boldsymbol{u}_t(\boldsymbol{x}_t)$ ▷ Incur reward $\Psi_{t,\alpha}(\boldsymbol{\theta}_t, \underline{\boldsymbol{x}_t})$ and loss $\Psi_{t,\alpha}(\underline{\boldsymbol{\theta}_t}, \boldsymbol{x}_t)$

5: $\boldsymbol{g}_{\mathcal{X},t} \in \partial_{\boldsymbol{x}} \Psi_{t,\alpha}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) = \sum_{i \in \mathcal{I}} \theta_{t,i} \partial_{\boldsymbol{x}} u_{t,i}$ ▷ Compute supergradient $\boldsymbol{g}_{\mathcal{X},t}$ at $\boldsymbol{x}_t$ of reward $\Psi_{t,\alpha}(\boldsymbol{\theta}_t, \cdot)$

6: $\boldsymbol{g}_{\Theta,t} = \nabla_{\boldsymbol{\theta}} \Psi_{t,\alpha}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) = \left(\left(-\theta_{t,i}\right)^{-1/\alpha} - \boldsymbol{u}_t(\boldsymbol{x}_t)\right)_{i \in \mathcal{I}}$ ▷ Compute gradient $\boldsymbol{g}_{\Theta,t}$ at $\boldsymbol{\theta}_t$ of loss $\Psi_{t,\alpha}(\cdot, \boldsymbol{x}_t)$

7: $\eta_{\mathcal{X},t} = \operatorname{diam}(\mathcal{X}) / \sqrt{\sum_{s=1}^{t} \left\|\boldsymbol{g}_{\mathcal{X},s}\right\|_2^2}; \eta_{\Theta,t} = \alpha u_{\min}^{-1-1/\alpha}/t$ ▷ Compute adaptive learning rates

8: $\boldsymbol{x}_{t+1} = \Pi_{\mathcal{X}}\left(\boldsymbol{x}_t + \eta_{\mathcal{X},t} \boldsymbol{g}_{\mathcal{X},t}\right); \boldsymbol{\theta}_{t+1} = \Pi_{\Theta}\left(\boldsymbol{\theta}_t - \eta_{\Theta,t} \boldsymbol{g}_{\Theta,t}\right)$ ▷ Compute a new allocation and dual decision

4.2 OHF Policy

Our policy employs a convex-concave function, composed of a convex conjugate term that tracks the global fairness metric in a dual (conjugate) space, and a weighted sum of utilities term that tracks the appropriate allocations in the primal space. This function is used by the policy to compute a gradient and a supergradient to adapt its internal state. In detail, we define the function $\Psi_\alpha : \Theta \times \mathcal{X} \to \mathbb{R}$ given by

$$\Psi_{t,\alpha}(\boldsymbol{\theta}, \boldsymbol{x}) \triangleq (-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \boldsymbol{u}_t(\boldsymbol{x}), \tag{13}$$

where $\Theta = \left[-1/u^{\alpha}_{\star,\min}, -1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{<0}$ is a subspace of the dual (conjugate) space, and $(-F_\alpha)^\star$ is the *convex conjugate* (see Definition 3 in Appendix) of $-F_\alpha$ given by for any $\boldsymbol{\theta} \in \Theta$

$$(-F_\alpha)^\star(\boldsymbol{\theta}) = \begin{cases} \sum_{i \in \mathcal{I}} \frac{\alpha(-\theta_i)^{1-1/\alpha} - 1}{1-\alpha} & \text{for } \alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}, \\ \sum_{i \in \mathcal{I}} -\log(-\theta_i) - 1 & \text{for } \alpha = 1. \end{cases} \tag{14}$$

The policy is summarized in Algorithm 1. The algorithm only requires as input: the set of eligible allocations $\mathcal{X}$, the α-fairness parameter in $\mathbb{R}^{\mathcal{I}}_{\geq 0}$, and the range $\left[u_{\star,\min}, u_{\star,\max}\right]$ of values of the average utility obtained by the optimal allocation (5), i.e., $\frac{1}{T} \sum_{t \in \mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_\star) \in \left[u_{\star,\min}, u_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{>0}$. We stress that the target time horizon T is *not* an input to the policy. The utility bounds $u^{\alpha}_{\star,\min}$ and $u^{\alpha}_{\star,\max}$ depend on the specific application. For example, for the virtualized caching system considered in Section 6, one could simply pick a small enough $\epsilon > 0$ as $u^{\alpha}_{\star,\min}$, and the maximum batch size weighted by the largest retrieval cost in the network as $u^{\alpha}_{\star,\max}$ (see Eq. (27)). However, if prior information is available to tighten this range, the performance of the algorithm is ameliorated, as reflected in the regret bound in Eq. (15).

The policy uses its input to initialize the dual (conjugate) subspace $\Theta = \left[-1/u^{\alpha}_{\star,\min}, -1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}}$, an allocation $\boldsymbol{x}_1 \in \mathcal{X}$, and a dual decision $\boldsymbol{\theta}_1 \in \Theta$ (lines 1–2 in Algorithm 1). At a given timeslot $t \in \mathcal{T}$, the allocation $\boldsymbol{x}_t$ is selected; then a vector-valued utility $\boldsymbol{u}_t(\cdot)$ is revealed and in turn $\Psi_{t,\alpha}(\cdot, \cdot)$ is revealed to the policy (line 4 in Algorithm 1). The supergradient $\boldsymbol{g}_{\mathcal{X},t}$ of $\Psi_{t,\alpha}(\boldsymbol{\theta}_t, \cdot)$ at point $\boldsymbol{x}_t \in \mathcal{X}$, and the gradient $\boldsymbol{g}_{\Theta,t}$ of $\Psi_{t,\alpha}(\cdot, \boldsymbol{x}_t)$ at point $\boldsymbol{\theta}_t \in \Theta$ are computed (lines 5–6 in Algorithm 1). The policy then finally performs an adaptation of its state variables $(\boldsymbol{x}_t, \boldsymbol{\theta}_t)$ through a descent step in the dual space and an ascent step in the primal space through online gradient descent (OGD) and online gradient ascent (OGA) policies,[3] respectively (line 8 in Algorithm 1). The

[3]Note that a different OCO policy can be used as long as it has a no-regret guarantee, e.g., online mirror descent (OMD), follow the regularized leader (FTRL), or follow the perturbed leader (FTPL) [34, 51]; moreover, one could even incorporate optimistic versions of such policies [59], to improve the regret rates when the controller has access to accurate predictions.

learning rates (step size) used are "self-confident" [7] as they depend on the experienced gradients. Such a learning rate schedule is compelling because it can adapt to the adversary and provides tighter regret guarantees for "easy" utility sequences; moreover, it allows attaining an *anytime* regret guarantee, i.e., a guarantee holding for any time horizon T. In particular, OHF policy in Algorithm 1 enjoys the following fairness regret guarantee.

Theorem 2. *Under assumptions (A1)–(A5), OHF policy in Algorithm 1 attains the following fairness regret guarantee:*

$$\mathfrak{R}_T(F_\alpha, \boldsymbol{\mathcal{A}}) \leq \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ \frac{1.5 \operatorname{diam}(\mathcal{X})}{T} \sqrt{\sum_{t\in\mathcal{T}} \left\|\boldsymbol{g}_{\mathcal{X},t}\right\|_2^2} + \sum_{t=1}^{T} \frac{\alpha \left\|\boldsymbol{g}_{\Theta,t}\right\|_2^2}{2u_{\star,\min}^{1+\frac{1}{\alpha}} T t} \right\} + O\left(\frac{\min\{\mathbb{V}_\mathcal{T}, \mathbb{W}_\mathcal{T}\}}{T}\right) \quad (15)$$

$$\leq \frac{1.5 \operatorname{diam}(\mathcal{X})\, L_\mathcal{X}}{u_{\star,\min}^{\alpha} \sqrt{T}} + \frac{\alpha L_\Theta^2 (\log(T)+1)}{u_{\star,\min}^{1+\frac{1}{\alpha}} T} + O\left(\frac{\min\{\mathbb{V}_\mathcal{T}, \mathbb{W}_\mathcal{T}\}}{T}\right) \quad (16)$$

$$= O\left(\frac{1}{\sqrt{T}} + \frac{\min\{\mathbb{V}_\mathcal{T}, \mathbb{W}_\mathcal{T}\}}{T}\right) = o(1). \quad (17)$$

The proof is provided in Appendix C. We prove that the fairness regret can be upper bounded with the time-averaged regrets of the primal policy operating over the set $\mathcal{X}$ and the dual policy operating over the set Θ, combined with an extra term that is upper bounded with $\min\{\mathbb{V}_\mathcal{T}, \mathbb{W}_\mathcal{T}\}$. Note that the fairness regret upper bound in Eq. (15) can be much tighter than the one in Eq. (16), because the gradients' norms can be smaller than their upper bound at a given timeslot $t \in \mathcal{T}$. Thanks to its "self-confident" learning schedule [7], which dynamically adapts to the observed utilities, our Algorithm 1 enjoys an *any-time* regret guarantee, i.e., it does not require the knowledge of the target time horizon T.

The result in Theorem 2 is tight, in the sense that no policy can have a fairness regret (8) with better dependency on the time horizon T. Formally,

Theorem 3. *Any policy $\boldsymbol{\mathcal{A}}$ incurs $\mathfrak{R}_T(F_\alpha, \boldsymbol{\mathcal{A}}) = \Omega\left(\frac{1}{\sqrt{T}}\right)$ fairness regret (8) for $\alpha \geq 0$.*

The proof can be found in Appendix D. We show that the lower bound on regret in online convex optimization [34] can be transferred to the fairness regret.

We discuss in Appendix H, the time-complexity of Algorithm 1 in the context of virtualized caching system application, presented in Section 6.

4.3 Adversarial Examples

In this section, we provide examples of adversaries satisfying Assumptions (A1)–(A5), with either $\mathbb{V}_\mathcal{T} = o(T)$ or $\mathbb{W}_\mathcal{T} = o(T)$, and of stochastic adversaries.

Example 2. (Adversaries satisfying $\mathbb{V}_\mathcal{T} = o(T)$) Consider an adversary selecting utilities such that

$$\boldsymbol{u}_t(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{\gamma}_t \odot \boldsymbol{p}_t(\boldsymbol{x}), \quad (18)$$

where $\boldsymbol{u} : \mathcal{X} \to \mathbb{R}^\mathcal{I}$ is a fixed utility, the time-dependent function $\boldsymbol{p}_t : \mathcal{X} \to \mathbb{R}^\mathcal{I}$ is an adversarially selected perturbation with $\left\|\boldsymbol{p}_t\right\|_\infty < \infty$, $\boldsymbol{\gamma}_t \in \mathbb{R}^\mathcal{I}$ quantifies the severity of the perturbations, and $\boldsymbol{\gamma}_t \odot \boldsymbol{p}_t(\boldsymbol{x}) = \left(\gamma_{t,i} p_{t,i}(\boldsymbol{x})\right)_{i\in\mathcal{I}}$ is the Hadamard product. The severity of the perturbations grows sublinearly in time T, i.e., $\sum_{t=1}^T \gamma_{t,i} = o(T)$ for every $i \in \mathcal{I}$. It is easy to check that, in this setting, it holds $\mathbb{V}_\mathcal{T} = o(T)$.

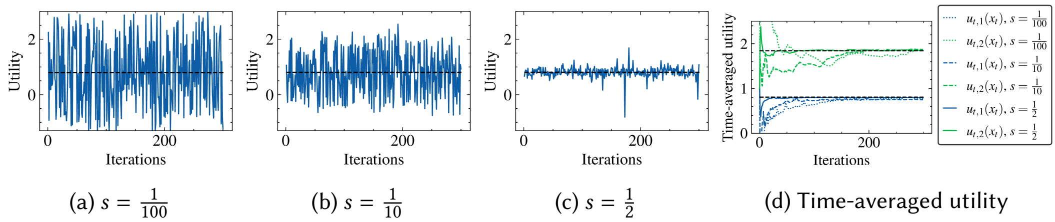


Fig. 2. Subfigures (a)–(c) provide the utilities of agent 2 for different values of perturbations' severity parameter $s \in \left\{\frac{1}{100}, \frac{1}{10}, \frac{1}{2}\right\}$ under the benchmark's allocation $x_\star$. Subfigure (d) provides the time-averaged utility of two agents. The dark dashed lines represent the utilities obtained by HF objective (5).

We provide a simple-yet-illustrative example of such an adversary. We take $\mathcal{X} = [0,1] \subset \mathbb{R}$, two agents $\mathcal{I} = \{1,2\}$, fixed utilities $\boldsymbol{u}(x) = \left(1-x^2, 1+x\right)$, adversarial perturbations $\boldsymbol{p}_t(x) = \left(a_{i,t} \cdot x\right)_{i \in \mathcal{I}}$ where $\boldsymbol{a}_t$ is selected uniformly at random from $[-1,1]^{\mathcal{I}}$ for every $t \in \mathcal{T}$. The perturbations' severity is selected as $\gamma_{\xi_{t,i},i} = t^{-s}$ where $\boldsymbol{\xi}_i : \mathcal{T} \to \mathcal{T}$ is a random permutation of the elements of $\mathcal{T}$ for $i \in \mathcal{I}$. The performance of Algorithm 1 is provided in Fig. 2. We observe that for larger values of s, corresponding to lower perturbation's severity, the policy provides faster the same utilities as the HF benchmark (5).

Example 3. (Adversaries satisfying $\mathbb{W}_{\mathcal{T}} = o(T)$) Consider a multiset $\mathcal{M}_t$ of utilities and an adversary that selects a utility $\boldsymbol{u}_t : \mathcal{X} \to \mathbb{R}^{\mathcal{I}}$ from it. The multiset is updated as follows: if $\mathcal{M}_t \setminus \{\boldsymbol{u}_t\} \neq \emptyset$, $\mathcal{M}_{t+1} = \mathcal{M}_t \setminus \{\boldsymbol{u}_t\}$, otherwise, $\mathcal{M}_t = \mathcal{M}_1$. In words, the adversary selects irrevocably elements (utilities) from the set $\mathcal{M}_1$, and, when all the elements are selected, the replenished $\mathcal{M}_1$ is offered again to the adversary. Consider, without loss of generality, a time horizon T divisible by $|\mathcal{M}_1|$ and the following decomposition for the period $\mathcal{T}$: $\{1,2,\ldots,|\mathcal{M}_1|\} \cup \{|\mathcal{M}_1|+1, |\mathcal{M}_1|+2, \ldots, 2\,|\mathcal{M}_1|\} \cup \cdots = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \cdots \cup \mathcal{T}_{T/|\mathcal{M}_1|}$. By construction, it holds for every $\boldsymbol{x} \in \mathcal{X}$

$$\sum_{i \in \mathcal{I}} \left| \sum_{t \in \mathcal{T}_k} \delta_{t,i}(\boldsymbol{x}) \right| = 0, \quad \forall k \in \{1,2,\ldots,T/|\mathcal{M}_1|\}, \tag{19}$$

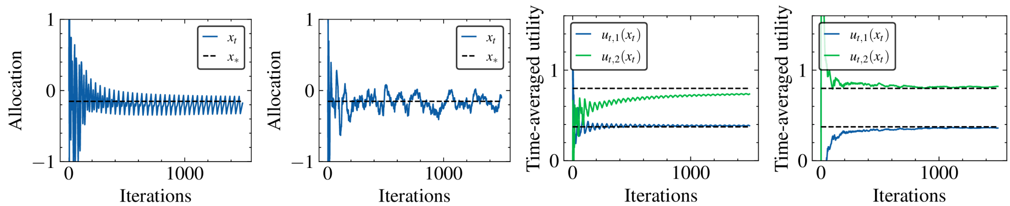
because when the multiset is fully consumed by the adversary, the average experienced utility is a fixed function. When $|\mathcal{M}_1| = \Theta\left(T^{\epsilon}\right)$ for $\epsilon \in [0,1/2)$ it holds $\sum_{k=1}^{T/|\mathcal{M}_1|} \frac{|\mathcal{T}_k|^2}{\sum_{k'<k}|\mathcal{T}_k|+1} = O\left(T^{2\epsilon}\right)$ (see Eq. (11)); thus, combined with Eq. (19) it holds $\mathbb{W}_{\mathcal{T}} = o(T)$. We provide a simple example of such an adversary. Consider $\mathcal{X} = [-1,1]$, two agents $\mathcal{I} = \{1,2\}$, and the initial multiset

$$\mathcal{M}_1 = \{\underbrace{(1-x, 1-(1-x)^2)}_{\text{repeated 10 times}}, \underbrace{(1-(1-x)^2, 1-4x)}_{\text{repeated 20 times}}, \underbrace{(1,-2x)}_{\text{repeated 10 times}}\}. \tag{20}$$

We have $|\mathcal{M}_1| = 40$ and hence $\mathbb{W}_{\mathcal{T}} = o(T)$. The performance of Algorithm 1 is provided in Fig. 3 under different choice patterns over $\mathcal{M}_1$. We observe that the cyclic choice of utilities is more harmful than the u.a.r. one as it leads to slower convergence. Nonetheless, under both settings, the policy asymptotically yields the same utilities as the HF benchmark (5).

Example 4. (Stochastic Adversary) Consider a scenario where $u_{t,i} : \mathcal{X} \to \mathbb{R}$ are drawn i.i.d. from an unknown distribution $\mathcal{D}_i$. Formally, the following corollary is obtained from Theorem 2.

Corollary 4. *When the utilities $u_{t,i} : \mathcal{X} \to \mathbb{R}$ are drawn i.i.d. from an unknown distribution $\mathcal{D}_i$ satisfying Assumptions (A1)–(A4), the policy OHF in Algorithm 1 attains the following expected*



(a) Allocations (cyclic) (b) Allocations (u.a.r.) (c) Time-averaged utilities (cyclic) (d) Time-averaged utilities (u.a.r.)

Fig. 3. Subfigures (a)–(b) provide the allocations of different agents of cyclic and u.a.r. choice of utilities over the set $\mathcal{M}_1$, respectively. Subfigures (c)–(d) provide the time-averaged utility of cyclic and u.a.r. choice of utilities over the set $\mathcal{M}_1$, respectively.

fairness regret guarantee:

$$\bar{\mathfrak{R}}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) \triangleq \sup_{\mathcal{D}_i, i \in \mathcal{I}} \left\{ \mathop{\mathbb{E}}_{\substack{u_{t,i} \sim \mathcal{D}_i \\ i \in \mathcal{I}, t \in \mathcal{T}}} \left[\max_{\boldsymbol{x} \in \mathcal{X}} F_\alpha \left(\frac{1}{T} \sum_{t \in \mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_t) \right) - F_\alpha \left(\frac{1}{T} \sum_{t \in \mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_t) \right) \right] \right\} = \mathcal{O}\left(\frac{1}{\sqrt{T}} \right). \quad (21)$$

Moreover, it holds with probability one: $\mathfrak{R}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) \leq 0$ *for* $T \to \infty$.

The proof is in Appendix E. The expected fairness regret guarantee follows from Theorem 2 and observing that $\mathbb{E}\left[\boldsymbol{\delta}_t(\boldsymbol{x})\right] = \mathbf{0}$ for any $t \in \mathcal{T}$ and $\boldsymbol{x} \in \mathcal{X}$. The high probability fairness regret guarantee for large T is obtained through Hoeffding's inequality paired with Eq. (10).

Note that we provide additional examples of adversaries, in the context of the application of our policy to a virtualized caching system, in Section 6.

5 Extensions

In this section, we first show that our algorithmic framework extends to cooperative bargaining settings, in particular Nash bargaining [55]. Secondly, we show that our framework also extends to the weighted α-fairness criterion.

5.1 Nash Bargaining

Nash bargaining solution (NBS), proposed in the seminal paper [55], is a fairness criterion for dispersing to a set of agents the utility of their cooperation. The solution guarantees that, whenever the agents cooperate, each agent achieves an individual performance that exceeds its performance when operating independently. This latter is also known as the disagreement point. NBS comes from the area of cooperative game theory, and it is self enforcing, i.e., the agents will agree to apply this solution without the need for an external authority to enforce compliance. NBS has been extensively applied in communication networks, e.g., to transmission power control [17], mobile Internet sharing among wireless users [36], content delivery in ISP-CDN partnerships [39], and cooperative caching in information-centric networks [75].

Nash bargaining can be incorporated through our fairness framework when $\alpha = 1$, and utilities as redefined for every $t \in \mathcal{T}$ as follows $\boldsymbol{u}'_t(\boldsymbol{x}) = \boldsymbol{u}_t(\boldsymbol{x}) - \boldsymbol{u}^d_t$ where u^d_i is the disagreement point of agent $i \in \mathcal{I}$. In particular, OHF provides the same guarantees. We also note that the dynamic model generalizes the NBS solution by allowing both the utilities and the disagreement points to change over time, while the benchmark is defined using (5) and $\alpha = 1$. Hence, the proposed OHF allows the agents to collaborate without knowing in advance the benefits of their cooperation nor their disagreement points, in a way that guarantees they will achieve the commonly agreed NBS at the end of the horizon T (asymptotically).

5.2 The $(\boldsymbol{w}, \alpha)$-Fairness

The weighted α-fairness or simply $(\boldsymbol{w}, \alpha)$-fairness with $\alpha \geq 0$ and $\boldsymbol{w} \in \Delta_{\mathcal{I}} \subset \mathbb{R}_{\geq 0}$, where $\Delta_{\mathcal{I}}$ is the probability simplex with support $\mathcal{I}$, is defined as [53]:

Definition 2. *A $(\boldsymbol{w}, \alpha)$-fairness function $F_{\boldsymbol{w},\alpha} : \mathcal{U} \to \mathbb{R}$ is parameterized by the inequality aversion parameter $\alpha \in \mathbb{R}_{\geq 0}$, weights $\boldsymbol{w} \in \Delta_{\mathcal{I}}$ and it is given by $F_{\boldsymbol{w},\alpha}(\boldsymbol{u}) \triangleq \sum_{i \in \mathcal{I}} w_i f_\alpha(u_i)$ for every $\boldsymbol{u} \in \mathcal{U}$. Note that $\mathcal{U} \subset \mathbb{R}^{\mathcal{I}}_{\geq 0}$ for $\alpha < 1$, and $\mathcal{U} \subset \mathbb{R}^{\mathcal{I}}_{> 0}$ for $\alpha \geq 1$.*

It is easy to check that our α-fairness framework captures the $(\boldsymbol{w}, \alpha)$-fairness by simply redefining the utilities incurred at time $t \in \mathcal{I}$ for agent $i \in \mathcal{I}$ as follows: $u'_{t,i}(\boldsymbol{x}) = w_i^{\frac{1}{1-\alpha}} u_{t,i}(\boldsymbol{x})$ for $\alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}$, otherwise $u'_{t,i}(\boldsymbol{x}) = \left(u_{t,i}(\boldsymbol{x})\right)^{w_i}$. Note that for $\alpha = 1$ and uniform weights, we recover the Nash bargaining setting discussed previously; otherwise, we recover asymmetric Nash bargaining in which the different weights correspond to the bargaining powers of players [33].

6 Application

In order to demonstrate the applicability of the proposed fairness framework, we target a representative resource management problem in virtualized caching systems where different caches cooperate by serving jointly the received content requests. This problem has been studied extensively in its static version, where the request rates for each content file are a priori known and the goal is to decide which files to store at each cache to maximize a fairness metric of cache hits across different caches, see for instance [48, 75]. We study the more realistic version of the problem where the request patterns are unknown. This online caching model has been recently studied as a learning problem in a series of papers [18, 45, 52, 56, 57, 67], yet none of them handles fairness metrics.

6.1 Multi-Agent Cache Networks

Cache network. We assume that time is slotted and the set of timeslots is denoted by $\mathcal{T} \triangleq \{1, 2, \ldots, T\}$. We consider a catalog of equally-sized files $\mathcal{F} \triangleq \{1, 2, \ldots, F\}$.[4] We model a cache network at timeslot $t \in \mathcal{T}$ as an undirected weighted graph $G_t(C, \mathcal{E})$, where $C \triangleq \{1, 2, \ldots, C\}$ is the set of caches, and $(c, c') \in \mathcal{E}$ denotes the link connecting cache c to c' with associated weight $w_{t,(c,c')} \in \mathbb{R}_{>0}$. Let $\mathcal{P}_{t,(c,c')} = \left\{c_1, c_2, \ldots, c_{|\mathcal{P}_{t,(c,c')}|}\right\} \in C^{|\mathcal{P}_{t,(c,c')}|}$ be the shortest path at timeslot $t \in \mathcal{T}$ from cache c to cache c' with associated weight $w^{\text{sp}}_{t,(c,c')} \triangleq \sum_{k=1}^{|\mathcal{P}_{t,(c,c')}|-1} w_{t,(c_k,c_{k+1})}$.

We assume for each file $f \in \mathcal{F}$ is permanently stored at a set $\Lambda_f(C) \subset C$ of designated repository servers. Moreover, each cache can store fractions of the file and fractions of the same file at different caches can be additively combined.[5] We denote by $x_{t,c,f} \in [0, 1]$ the fraction of file $f \in \mathcal{F}$ stored at cache $c \in C$ at timeslot $t \in \mathcal{T}$. The state of cache $c \in C$ is given by $\boldsymbol{x}_{t,c}$ drawn from the set

$$\mathcal{X}_c \triangleq \left\{ \boldsymbol{x} \in [0, 1]^{\mathcal{F}} : \sum_{f \in \mathcal{F}} x_f \leq k_c, x_f \geq \mathbb{1}\left(c \in \Lambda_f(C)\right), \forall f \in \mathcal{F} \right\}, \tag{22}$$

[4]Note that we assume equally-sized files to streamline the presentation. Our model supports unequally-sized files by replacing the cardinality constraint in Eq. (22) with a knapsack constraint and the set $\mathcal{X}_c$ (defined in (22)) remains convex.
[5] This is a common assumption [57, 65], which models situations where each file can be split in a large number of small chunks and each cache can store random linear combinations of such chunks. Guarantees for this fractional setting can be readily transferred to an integral setting through randomized rounding techniques [16, 35, 56, 68].

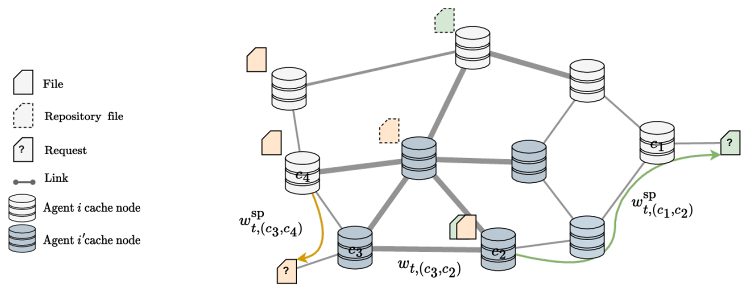


Fig. 4. System model: a network comprised of a set of caching nodes C. A request arrives at a cache node $c \in C$, it can be partially served locally, and if needed, forwarded along the shortest retrieval path to another node to retrieve the remaining part of the file; a utility is incurred by the cache owner $i \in \mathcal{I}$. A set of permanently allocated files are spread across the network guaranteeing requests can always be served.

where $k_c \in \mathbb{N}$ is the capacity of cache $c \in C$, and $\mathbb{1}(\chi) \in \{0, 1\}$ is the indicator function set to 1 when condition χ is true. Thus, the state of the cache network belongs to $\mathcal{X} \triangleq \bigtimes_{c \in C} \mathcal{X}_c$. The system model is summarized in Fig. 4, and it is aligned with many recent papers focusing on learning for caching [35, 56, 57, 68].

Requests. We denote by $r_{t,c,f} \in \mathbb{N} \cup \{0\}$ the number of requests for file $f \in \mathcal{F}$ submitted by users associated to cache $c \in C$, during slot $t \in \mathcal{T}$. The request batch arriving at timeslot $t \in \mathcal{T}$ is denoted by $\boldsymbol{r}_t = \left(r_{t,c,f}\right)_{(c,f) \in C \times \mathcal{F}}$ and belongs to the set

$$\mathcal{R}_t \triangleq \left\{ \boldsymbol{r} \in (\mathbb{N} \cup \{0\})^{C \times \mathcal{F}} : \sum_{c \in C} \sum_{f \in \mathcal{F}} r_{c,f} \leq R_t \right\},$$

where $R_t \in \mathbb{N}$ is the total number of requests (potentially) arriving at the system at timeslot $t \in \mathcal{T}$.

Caching gain. We consider an agent $i \in \mathcal{I}$ holds a set of caches $\Gamma_i(C) \subset C$, and $\dot{\bigcup}_{i \in \mathcal{I}} \Gamma_i(C) = C$. Hence, the allocation set of agent i is given by $\mathcal{X}_i = \bigtimes_{c \in \Gamma_i(C)} \mathcal{X}_c$. Requests arriving at cache $c \in C$ can be partially served locally, and if needed, forwarded along the shortest path to a nearby cache $c' \in C$ storing the file, incurring a retrieval cost $w^{\text{sp}}_{t,(c,c')}$. Let $\phi_{t,i,c} \triangleq \arg\min_{c' \in \Lambda_i(C)} \left\{ w^{\text{sp}}_{t,(c,c')} \right\}$ and $\Phi_{t,i,c} : \left\{1, 2, \ldots, \phi_{t,i,c}\right\} \subset C \to C$ be a map providing a retrieval cost ordering for every $c \in \left\{1, 2, \ldots, \phi_{t,i,c}\right\}$, $t \in \mathcal{T}$, and $i \in \mathcal{I}$, i.e.,

$$w^{\text{sp}}_{t,(c,\Phi_{t,i,c}(\phi_{t,i,c}))} = \min \left\{ w^{\text{sp}}_{t,(c,c')} : c' \in \Lambda_f(C) \right\} \geq \cdots \geq w^{\text{sp}}_{t,(c,\Phi_{t,i,c}(2)))} \geq w^{\text{sp}}_{t,(c,\Phi_{t,i,c}(1)))} = 0. \tag{23}$$

When a request batch $\boldsymbol{r}_t \in \mathcal{R}_t$ arrives at timeslot $t \in \mathcal{T}$, agent $i \in \mathcal{I}$ incurs the following cost:

$$\text{cost}_{t,i}(\boldsymbol{x}) \triangleq \sum_{c \in \Gamma_i(C)} \sum_{f \in \mathcal{F}} r_{t,c,f} \sum_{k=1}^{\phi_{t,i,c}-1} \left(w^{\text{sp}}_{t,(c,\Phi_{t,i,c}(k+1))} - w^{\text{sp}}_{t,(c,\Phi_{t,i,c}(k))} \right) \left(1 - \min \left\{ 1, \sum_{k'=1}^{k} x_{\Phi_{t,i,c}(k'),f} \right\} \right).$$

This can be interpreted as a QoS cost paid by a user for the additional delay to retrieve part of the file from another cache, or it can represent the load on the network to provide the missing file. Note that by construction, the maximum cost is achieved for a network state, where all the caches are empty except for the repository allocations; formally, such state is given by $\boldsymbol{x}_0 \triangleq$

$\left(\mathbb{1}\left(c \in \Lambda_f(C)\right)\right)_{(c,f)\in C\times\mathcal{F}} \in \mathcal{X}$, and the cost of the agent at this state is given by

$$\mathrm{cost}_{t,i}(\boldsymbol{x}_0) = \sum_{c\in\Gamma_i(C)} \sum_{f\in\mathcal{F}} r_{t,c,f} \min\left\{w^{\mathrm{sp}}_{t,(c,c')} : c' \in \Lambda_f(C)\right\} \tag{24}$$

$$= \sum_{c\in\Gamma_i(C)} \sum_{f\in\mathcal{F}} r_{t,c,f} \sum_{k=1}^{\phi_{t,i,c}-1} \left(w^{\mathrm{sp}}_{t,(c,\Phi_{t,i,c}(k+1))} - w^{\mathrm{sp}}_{t,(c,\Phi_{t,i,c}(k))}\right), \tag{25}$$

We can define the caching utility at timeslot $t \in \mathcal{T}$ as the cost reduction due to caching as:

$$u_{t,i}(\boldsymbol{x}) \triangleq \mathrm{cost}_{t,i}(\boldsymbol{x}_0) - \mathrm{cost}_{t,i}(\boldsymbol{x}) \tag{26}$$

$$= \sum_{c\in\Gamma_i(C)} \sum_{f\in\mathcal{F}} r_{t,c,f} \sum_{k=1}^{\phi_{t,i,c}-1} \left(w^{\mathrm{sp}}_{t,(c,\Phi_{t,i,c}(k+1))} - w^{\mathrm{sp}}_{t,(c,\Phi_{t,i,c}(k))}\right) \min\left\{1, \sum_{k'=1}^{k} x_{\Phi_{t,i,c}(k'),f}\right\}. \tag{27}$$

The caching utility is a weighted sum of concave functions with positive weights, and thus concave in $\boldsymbol{x} \in \mathcal{X}$. It is straightforward to check that this problem always satisfies Assumptions (A1)–(A4). The request batches and the time-varying retrieval costs determine whether Assumption (A5) holds. For example, this is the case when request batches are drawn i.i.d. from a fixed unknown distribution (see Example 4).

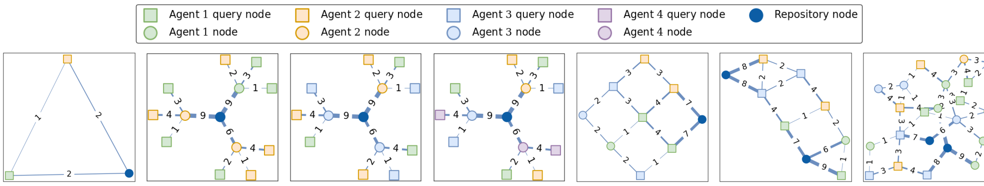
6.2 Results

Below we describe the experimental setup[6] of the multi-agent cache networks problem, the request traces, and competing policies. Our results are summarized as follows:

(1) Under stationary requests and small batch sizes (leading to large utility deviations from one timeslot to another), OHF achieves the same time-averaged utilities as the offline benchmark, whereas OSF, a counterpart policy to OHF targeting slot-fairness (4), diverges and is unable to reach the Pareto front.
(2) In the Nash bargaining scenario, OHF achieves the NBS in all cases, while OSF fails when the disagreement points are exigent, i.e., an agent can guarantee itself a high utility.
(3) Widely used LFU and LRU might perform arbitrarily bad w.r.t. fairness, and not even achieve any point in the Pareto front (hence, they are not only unfair, but also inefficient).
(4) Fairness comes at a higher price when α is increased or the number of agents is increased. This observation on the price of fairness provides experimental evidence for previous work [14].
(5) OHF is robust to different network topologies and is able to obtain time-averaged utilities that match the offline benchmark.
(6) Under non-stationary requests, OHF policy achieves the same time-averaged utilities as the offline benchmark, whereas OSF can perform arbitrarily bad providing allocations that are both unfair and inefficient

General Setup. We consider three synthetic network topologies (Cycle, Tree, and Grid), and two real network topologies (Abilene and GEANT). A visualization of the network topologies is provided in Figure 5. The specifications of the network topologies used across the experiments are provided in Table 2 in the Appendix. A repository node permanently stores the entire catalog of files. The retrieval costs along the edges are sampled u.a.r. from $\{1, 2, \ldots, 5\}$, except for edges directly connected to a repository node which are sampled u.a.r. from $\{6, 7, \ldots, 10\}$. All the retrieval costs remain fixed for every $t \in \mathcal{T}$. The capacity of each cache is sampled u.a.r. from $\{1, 2, \ldots, 5\}$, but for the Cycle topology in which each cache has capacity 5. An agent $i \in \mathcal{I}$ has a set of query

[6]Our code is publicly available at https://github.com/tareq-si-salem/Online-Multi-Agent-Cache-Networks



(a) Cycle (b) Tree-1 (c) Tree-2 (d) Tree-3 (e) Grid (f) Abilene (g) GEANT

Fig. 5. Network topologies used in experiments.

nodes denoted by $Q_i \subset \Gamma_i(C)$, and a query node can generate a batch of requests from a catalog with $|\mathcal{F}| = 20$ files. Unless otherwise said, we consider $u_{\star,\min} = 0.1$ and $u_{\star,\max} = 1.0$. The fairness benchmark refers to the maximizer of the HF objective (5), and the utilitarian benchmark refers to the maximizer of HF objective (5) for $\alpha = 0$.

Traces. Each query node generates requests according to the following:

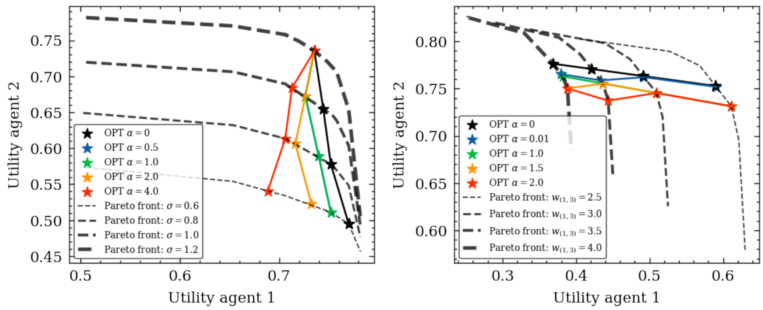
- Stationary trace (parameters: σ, R, T, F). Requests are sampled i.i.d. from a Zipf distribution with exponent $\sigma \in \mathbb{R}_{\geq 0}$ from a catalog of files of size F. The requests are grouped into batches of size $|\mathcal{R}_t| = R, \forall t \in \mathcal{T}$.
- Non-Stationary trace (parameters: σ, R, T, F, D). Similarly, requests are sampled i.i.d. from a catalog of F files according to a Zipf distribution with exponent $\sigma \in \mathbb{R}_{\geq 0}$. Every D requests, the popularity distribution is modified in the following fashion: file $f \in \mathcal{F} = \{1, 2, \ldots, F\}$ assumes the popularity of file $f' = (f + F/2) \mod F$ (F is even). The requests are grouped into batches of size $|\mathcal{R}_t| = R, \forall t \in \mathcal{T}$.

The stationary trace corresponds to the stochastic adversary in Example 4, and the non-stationary trace corresponds to a stochastic adversary with perturbations satisfying the partitioned-severity condition in Eq. (10). Two sampled traces are depicted in Figure 13 in the Appendix. Unless otherwise said, query nodes generate *Stationary* traces and $\sigma = 1.2$, $T = 10^4$, $R = 50$, and $D = 50$.

Policies. We implement the following policies and use them as comparison benchmarks for OHF.

- The classic LRU and LFU policies. A request is routed to the cache with minimal retrieval cost among those that store the requested file and this cache provides the content and updates its state corresponding to a hit. Moreover, all caches with a lower retrieval cost update their state as if a miss occurred locally. This corresponds to the popular path replication algorithm [21, 35], equipped with LRU or LFU, adapted to our setting.
- Online slot-fairness (OSF) policy. This policy is the slot-fairness (4) counterpart of OHF. It is obtained by configuring Algorithm 1 with dual (conjugate) subspace $\Theta = \{(-1)_{i \in \mathcal{I}}\}$ (i.e., taking $\alpha \to 0$), which makes ineffective the dual policy in Algorithm 1. The revealed utilities at timeslot $t \in \mathcal{T}$ are the α-fairness transformed utilities $\boldsymbol{u}'_t(\cdot) = (f_\alpha(u_{t,i}(\cdot)))_{i \in \mathcal{I}}$. The primal allocations are still determined by the same self-confident learning rates' schedule as OHF for a fair comparison. The resulting policy is a no-regret policy (see Lemma 3 in Appendix) w.r.t. the slot-fairness benchmark (4) for some $\alpha \in \mathbb{R}_{\geq 0}$.

Static analysis of symmetry-breaking parameters. We start with a numerical investigation of the potential caching gains, and how these are affected by the fairness parameter α. In Figure 6, we consider the Cycle topology and different values of $\alpha \in [0, 2]$. We show the impact on the fairness benchmark of varying the request patterns ($\sigma \in \{0.6, 0.8, 1.0, 1.2\}$) for agent 2 under the *Stationary* trace in Fig. 6 (a), and of varying the retrieval costs between agent 1's cache and the repository



(a) (b)

Fig. 6. Pareto front and fairness benchmark's utilities for different values of $\alpha \in [0, 2]$ under different request patterns (a) ($\sigma \in \{0.6, 0.8, 1.0, 1.2\}$) for agent 2, and different retrieval costs (b) between agent 1's cache and the repository ($w_{(1,3)} \in [2.5, 4.0]$).

($w_{(1,3)} \in [2.5, 4]$). In Figure 6 (a), we observe decreasing the skewness of the popularity distribution decreases the utility of agent 2 as reflected by the downward shift of the Pareto front. We note that, as far as the file popularity distribution at agent 2 is close to the one at agent 1 ($\sigma = 1.2$), different values of alpha still provide similar utilities. However, in highly asymmetric scenarios, different values of α lead to clearly distinct utilities for each agent. We also note that higher values of α guarantees higher fairness by that increasing the utility of agent 2. Similarly, in Figure 6 (b), we observe increasing the retrieval cost for agent 1 decreases the utility achieved by the same agent, as reflected by the leftward shift of the Pareto front; moreover, increasing the retrieval costs (higher asymmetry) highlights the difference between different values of α.

Online analysis of symmetry-breaking parameters. In Figure 7, we consider the Cycle topology, and different values of $\alpha \in \{0, 1, 2\}$. In Figure 7 (a)–(b) we consider the retrieval cost $w_{(1,3)} = 3.5$ between agent 1's cache node and the repository node. In Figure 7 (c)–(d) query node of agent 1 generates *Stationary* trace ($\sigma = 1.2$) and query node of agent 2 generates *Stationary* trace ($\sigma = 0.6$). We consider two fixed request batch sizes $R \in \{1, 50\}$.

In Figures 7 (a) and (c) (for batch size $R = 1$) OHF approaches the fairness benchmark's utilities for different values of α, but OSF diverges for values of $\alpha \neq 0$. For increased request batch size $R = 50$, OHF and OSF exhibit similar behavior. This is expected under stationary utilities; increasing the batch size reduces the variability in the incurred utilities at every timeslot, and the horizon-fairness and slot-fairness objectives become closer yielding similar allocations. Note that this observation implies that OSF is only capable to converge for utilities with low variability, which is far from realistic scenarios. LFU policy outperforms LRU and both policies do not approach the Pareto front; thus, the allocations selected by such policies are inefficient and unfair.

Nash bargaining. In Figure 8, we consider the Cycle topology and $\alpha = 1$. We select different disagreement utilities for agent 2 in $\{0.0, 0.5, 0.7, 0.75\}$, i.e., different utility values agent 2 expects to guarantee itself even in the absence of cooperation. Note how higher values of disagreement utilities lead to higher utilities for agent 2 at the fairness benchmark. We select $u_{\star,\min} = 0.01$.

For a small batch size ($R = 1$), OHF approaches the same utilities achieved by the fairness benchmark for different disagreement points, whereas OSF fails to approach the Pareto front. Similarly, for a larger batch size $R = 50$, OHF approaches the fairness benchmark for different disagreement points, but the Pareto front is reached faster than with a batch size $R = 1$. OSF

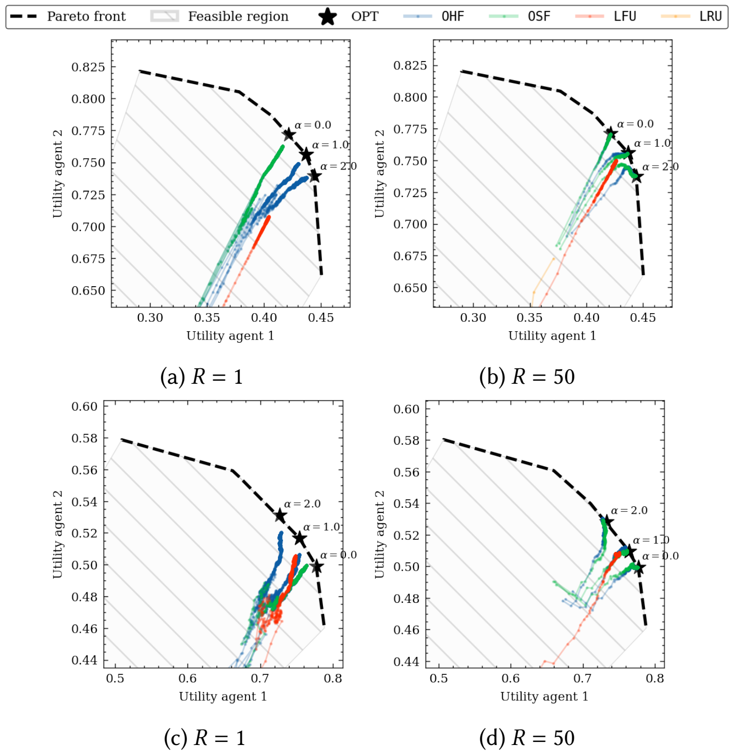
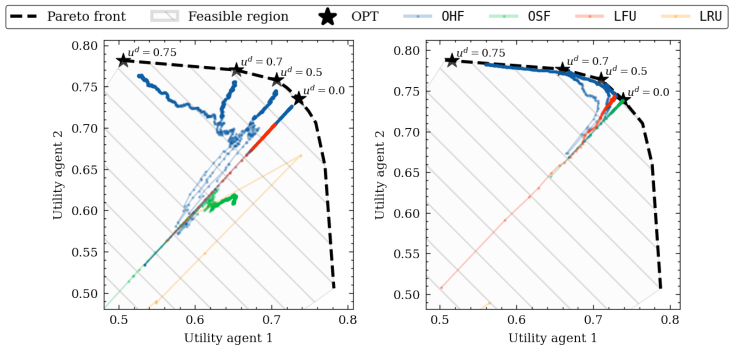


Fig. 7. Time-averaged utilities of policies OHF, OSF, LRU, and LFU under Cycle topology. Subfigures (a)–(b) are obtained under retrieval cost $w_{(1,3)} = 3.5$ for agent 1's query node. Subfigures (c)–(d) are obtained when agent 2's query node generates *Stationary trace* ($\sigma = 0.6$). Markers correspond to iterations in $\{100, 200, \ldots, 10^4\}$.

diverges for non-zero disagreement points when $R = 50$, because the allocation selected for some agent $i \in \mathcal{I}$ can be smaller than its disagreement utility (i.e., $u_{t,i}(\boldsymbol{x}_t) - u_{t,i} < 0$), while the α-fairness function is only defined for positive arguments.

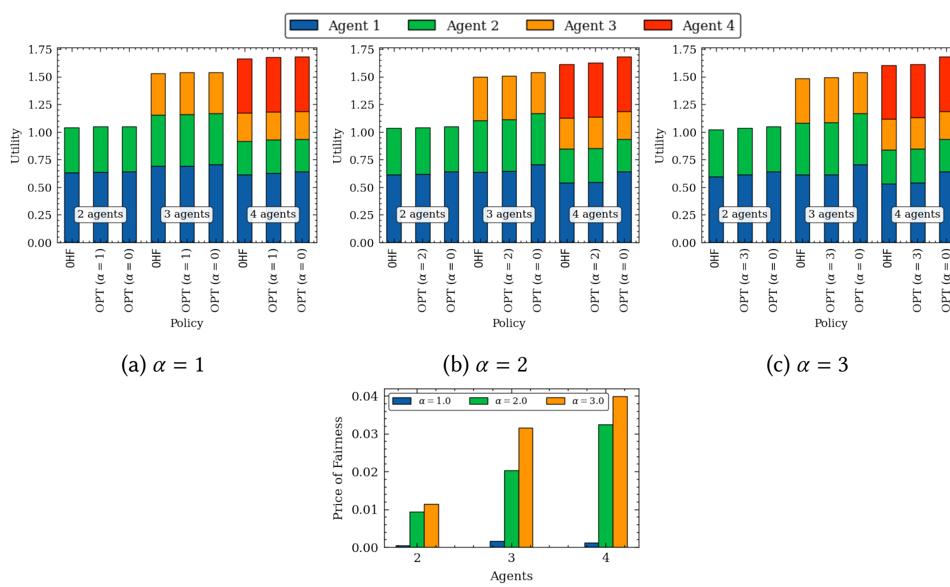
Impact of agents on the price of fairness. In Figures 9 and 10, we consider the Tree 1–3 topology, $\alpha \in \{1, 2, 3\}$, and $|\mathcal{I}| \in \{2, 3, 4\}$. Agents' query nodes generate *Stationary* trace ($\sigma \in \{1.2, 0.8, 0.6\}$). In Figures 9 (a)–(c), we observe for increasing the number of agents, the division of utilities differs between the fairness benchmark and utilitarian benchmark; moreover, this difference is more evident for larger values of α. Figure 9 (d) provides the price of fairness, and we observe the price of fairness increases with the number of agents and α. Nonetheless, under the different settings the price of fairness remains below 4%, i.e., we experience at most a 4% drop in the social welfare to provide fair utility distribution across the different agents. Figure 10 gives the time-averaged utilities obtained by running OHF for $\alpha = 2$. We observe the utilities obtained by OHF quickly converge to the same utilities obtained by the fairness benchmark. In this figure, we also highlight the difference between the utilities achieved by the fairness benchmark and utilitarian benchmark, is reflected by the increasing utility gap for a higher number of participating agents.

Different network topologies. In Figure 11 (a), we consider the network topologies Tree, Grid, Abilene, GEANT under *Stationary* trace ($\sigma \in \{0.6, 1.0, 1.2\}$) and $\alpha = 3$. OHF achieves the same



(a) $R = 1$ (b) $R = 50$

Fig. 8. Time-averaged utilities obtained for policies OHF, OSF, LRU, and LFU for batch sizes (a) $R = 1$ and (b) $R = 100$, under CYCLE network topology. Markers correspond to iterations in $\{100, 200, \ldots, 10^4\}$.



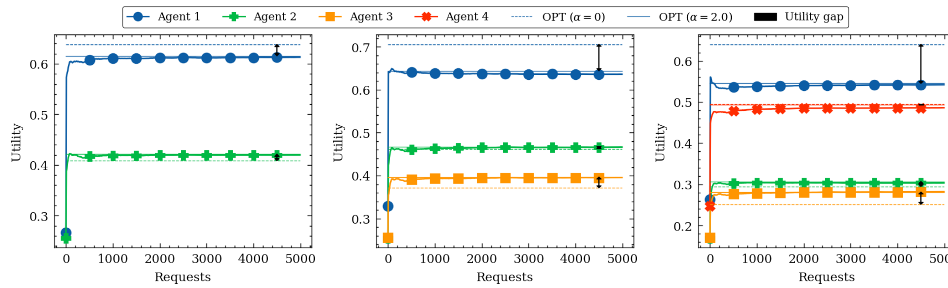
(a) $\alpha = 1$ (b) $\alpha = 2$ (c) $\alpha = 3$

(d) PoF

Fig. 9. Subfigures (a)–(c) provide the average utility for different agents obtained by OHF, fairness benchmark (OPT for $\alpha \neq 0$), and utilitarian benchmark (OPT for $\alpha = 0$); and Subfigure (d) provides the PoF for $\alpha \in \{0, 1, 2, 3\}$ under an increasing number of agents in $\{2, 3, 4\}$ and TREE 1–3 network topology.

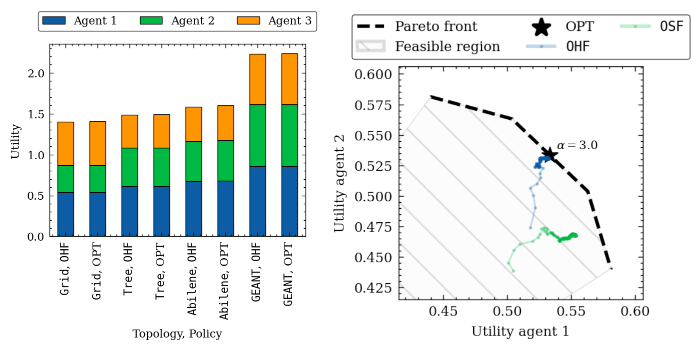
utilities as the fairness benchmark across the different topologies. Note that for larger network topologies agents achieve a higher utility because there are more resources available.

Impact of non-stationarity. In Figure 11 (b), we consider the CYCLE topology and $\alpha = 3$. The query node of agent 1 generates *Non-Stationary* trace, while the query node of agent 2 generates a shuffled *Non-Stationary* trace, i.e., we remove the non-stationarity from the trace for agent 2 while preserving the overall popularity of the requests. Therefore, on average the agents are symmetric



(a) (b) (c)

Fig. 10. Subfigures (a)–(c) provide the time-averaged utility across different agents obtained by OHF policy and OPT for $\alpha = 2$ under an increasing number of agents in $\{2, 3, 4\}$ and Tree 1–3 network topology.



(a) (b)

Fig. 11. Subfigure (a) provides the average utility of OHF and fairness benchmark under network topologies Tree, Grid, Abilene, GEANT, *Stationary* trace ($\sigma \in \{0.6, 0.8, 1.2\}$), and $\alpha = 3$. Subfigure (b) provides the time-averaged utilities obtained for OHF, OSF and batch size $R = 50, t \in \mathcal{T}$, under network topology Tree (a) and *Non-stationary* trace. The markers represent the iterations in the set $\left\{100, 200, \ldots, 10^4\right\}$.

and experience the same utilities. We observe in Figure 11 (b) that indeed this is the case for OHF policy; however, because OSF aims to insure fairness across the different timeslots the agents are not considered symmetric and the average utilities deviate from the Pareto front (not efficient). OSF favors agent 1 by increasing its utility by 20% compared to the utility of agent 1.

7 Conclusion and Future Work

In this work, we proposed a novel OHF policy that achieves horizon-fairness in dynamic resource allocation problems. We demonstrated the applicability of this policy in virtualized caching systems where different agents can cooperate to increase their caching gain. Our work paves the road for several interesting next steps. A future research direction is to consider decentralized versions of the policy under which each agent selects an allocation with limited information exchange across agents. For the application to virtualized caching systems, the message exchange techniques in [35, 45] to estimate subgradients can be exploited. Another important future research direction is to bridge the horizon-fairness and slot-fairness criteria to target applications where the agents are interested in ensuring fairness within a target time window. We observed that OHF can encapsulate the two criteria, however, it remains an open question whether a policy can smoothly transition between them. A final interesting research direction is to consider a limited feedback scenario where only part of the utility is revealed to the agents (e.g., bandit feedback). Our policy could be extended to this setting through gradient estimation techniques [34].

8 Acknowledgments

This publication has emanated from research conducted with the financial support of the European Commission through Grant No. 101017109 (DAEMON). This research was supported in part by the French Government through the “Plan de Relance” and “Programme d’investissements d’avenir”.

References

[1] Shipra Agrawal and Nikhil R. Devanur. 2014. Bandits with Concave Rewards and Convex Knapsacks. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation* (Palo Alto, California, USA) *(EC '14)*. Association for Computing Machinery, New York, NY, USA, 989–1006.

[2] Omid Alipourfard, Hongqiang Harry Liu, Jianshu Chen, Shivaram Venkataraman, Minlan Yu, and Ming Zhang. 2017. Cherrypick: adaptively unearthing the best cloud configurations for big data analytics. In *NSDI'17: Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation*. USENIX Association, USA, 469–482.

[3] Eitan Altman, Konstantin Avrachenkov, and Andrey Garnaev. 2008. Generalized α-fair resource allocation in wireless networks. In *2008 47th IEEE Conference on Decision and Control*. IEEE, 2414–2419.

[4] Eitan Altman, Konstantin Avrachenkov, and Sreenath Ramanath. 2012. Multiscale Fairness and its Application to Resource Allocation in Wireless Networks. *Computer Communications* 35, 7 (2012), 820–828.

[5] Daron Anderson, George Iosifidis, and Douglas J Leith. 2022. Lazy Lagrangians with Predictions for Online Learning. *arXiv preprint arXiv:2201.02890* (2022).

[6] Anthony B. Atkinson. 1970. On the Measurement of Inequality. *Journal of Economic Theory* 2, 3 (1970), 244–263.

[7] Peter Auer, Nicolò Cesa-Bianchi, and Claudio Gentile. 2002. Adaptive and Self-Confident On-Line Learning Algorithms. *J. Comput. System Sci.* 64, 1 (2002), 48–75.

[8] Jackie Baek and Vivek Farias. 2021. Fair exploration via axiomatic bargaining. *Advances in Neural Information Processing Systems* 34 (2021), 22034–22045.

[9] Santiago R Balseiro, Haihao Lu, and Vahab Mirrokni. 2022. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research* (2022).

[10] Siddhartha Banerjee, Vasilis Gkatzelis, Artur Gorokh, and Billy Jin. 2022. Online nash social welfare maximization with predictions. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1–19.

[11] MohammadHossein Bateni, Yiwei Chen, Dragos Florin Ciocan, and Vahab Mirrokni. 2022. Fair resource allocation in a volatile marketplace. *Operations Research* 70, 1 (2022), 288–308.

[12] Gerdus Benade, Aleksandr M Kazachkov, Ariel D Procaccia, and Christos-Alexandros Psomas. 2018. How to make envy vanish over time. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. 593–610.

[13] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. 2011. The Price of Fairness. *Operations Research* 59, 1 (2011).

[14] Dimitris Bertsimas, Vivek F Farias, and Nikolaos Trichakis. 2012. On The Efficiency-Fairness Trade-Off. *Management Science* 58, 12 (2012), 2234–2250.

[15] David Blackwell. 1956. An Analog of the Minmax Theorem for Vector Payoffs. *Pacific J. Math.* 6, 1 (1956), 1–8.

[16] B. Blaszczyszyn and A. Giovanidis. 2015. Optimal Geographic Caching In Cellular Networks. In *ICC*. 3358–3363.

[17] Holger Boche and Martin Schubert. 2011. A Generalization of Nash Bargaining and Proportional Fairness to Log-Convex Utility Sets With Power Constraints. *IEEE Transactions on Information Theory* 57, 6 (2011), 3390–3404.

[18] Archana Bura, Desik Rengarajan, Dileep Kalathil, Srinivas Shakkottai, and Jean-Francois Chamberland. 2021. Learning to Cache and Caching to Learn: Regret Analysis of Caching Algorithms. *IEEE/ACM Transactions on Networking* 30, 1 (2021), 18–31.

[19] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang. 2013. Multiresource Allocation: Fairness-Efficiency Tradeoffs in a Unifying Framework. *IEEE/ACM Trans. on Networking* 21, 6 (2013), 1785–1798.

[20] Semih Cayci, Swati Gupta, and Atilla Eryilmaz. 2020. Group-Fair Online Allocation in Continuous Time. *Advances in Neural Information Processing Systems* 33 (2020), 13750–13761.

[21] Edith Cohen and Scott Shenker. 2002. Replication Strategies in Unstructured Peer-to-Peer Networks. In *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (Pittsburgh, Pennsylvania, USA) *(SIGCOMM '02)*. Association for Computing Machinery, New York, NY, USA, 177–190.

[22] Steven Diamond and Stephen Boyd. 2016. CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research* 17, 1 (jan 2016), 2909–2913.

[23] John C Duchi, Alekh Agarwal, Mikael Johansson, and Michael I Jordan. 2012. Ergodic mirror descent. *SIAM Journal on Optimization* 22, 4 (2012), 1549–1578.

[24] Francis Ysidro Edgeworth et al. 1881. Mathematical Psychics. *History of Economic Thought Books* (1881).

[25] Edmund Eisenberg and David Gale. 1959. Consensus of subjective probabilities: The pari-mutuel method. *The Annals of Mathematical Statistics* 30, 1 (1959), 165–168.

[26] Salah-Eddine Elayoubi and James Roberts. 2015. Performance and cost effectiveness of caching in mobile access networks. In *Proceedings of the 2nd ACM Conference on Information-Centric Networking*. 79–88.

[27] Hossein Esfandiari, Nitish Korula, and Vahab Mirrokni. 2015. Online allocation with traffic spikes: Mixing adversarial and stochastic models. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. 169–186.

[28] Eyal Even-Dar, Robert Kleinberg, Shie Mannor, and Yishay Mansour. 2009. Online Learning with Global Cost Functions. In *22nd Annual Conference on Learning Theory, COLT*.

[29] F.P. Kelly, A. Maulloo, and D. Tan. 1998. Rate Control for Communication Networks: Shadow Prices, Proportional Fairness, and Stability. *J. Oper. Res. Soc.* 49, 3 (1998), 237–252.

[30] G. S. Paschos, E. Bastug, I. Land, G. Caire, M. Debbah. 2016. Wireless Caching: Technical Misconceptions and Business Barriers. *IEEE Commun. Mag.* 54, 8 (2016), 16–22.

[31] Leonidas Georgiadis, George Iosifidis, and Leandros Tassiulas. 2015. Exchange of services in networks: competition, cooperation, and fairness. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*. 43–56.

[32] Swati Gupta and Vijay Kamble. 2021. Individual Fairness in Hindsight. *J. Mach. Learn. Res.* 22, 144 (2021), 1–35.

[33] John C Harsanyi and Reinhard Selten. 1972. A generalized Nash solution for two-person bargaining games with incomplete information. *Management science* 18, 5-part-2 (1972), 80–106.

[34] Elad Hazan. 2016. Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization* 2, 3–4 (Aug. 2016), 157–325.

[35] Stratis Ioannidis and Edmund Yeh. 2016. Adaptive Caching Networks with Optimality Guarantees. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* (Antibes Juan-les-Pins, France) *(SIGMETRICS '16)*. Association for Computing Machinery, New York, NY, USA, 113–124.

[36] George Iosifidis, Lin Gao, Jianwei Huang, and Leandros Tassiulas. 2017. Efficient and Fair Collaborative Mobile Internet Access. *IEEE/ACM Transactions on Networking* 25, 3 (2017), 1386–1400.

[37] J. G. Andrews, et al. 2014. What Will 5G be? *IEEE Journal on Selected Areas in Communications* 32, 6 (2014), 1065–1082.

[38] Devansh Jalota and Yinyu Ye. 2022. Online Learning in Fisher Markets with Unknown Agent Preferences. *arXiv preprint arXiv:2205.00825* (2022).

[39] Wenjie Jiang, Rui Zhang-Shen, Jennifer Rexford, and Mung Chiang. 2009. Cooperative Content Distribution and Traffic Engineering in an ISP Network. *SIGMETRICS Performance Evaluation Review* 37, 1 (jun 2009), 239–250.

[40] Jose A. Ayala-Romero, Andres Garcia-Saavedra, Xavier Costa-Pérez, George Iosifidis. 2021. EdgeBOL: Automating Energy-savings for Mobile Edge AI. In *ACM CoNEXT*. 397–410.

[41] Ehud Kalai, Meir Smorodinsky, et al. 1975. Other Solutions to Nash's Bargaining Problem. *Econometrica* 43, 3 (1975), 513–518.

[42] Yasushi Kawase and Hanna Sumita. 2021. Online Max-min Fair Allocation. *arXiv preprint arXiv:2111.07235* (2021).

[43] Pawan Kumar and Rakesh Kumar. 2019. Issues and Challenges of Load Balancing Techniques in Cloud Computing: A Survey. 51, 6, Article 120 (feb 2019), 35 pages.

[44] Mathieu Leconte, Georgios Paschos, Lazaros Gkatzikis, Moez Draief, Spyridon Vassilaras, and Symeon Chouvardas. 2016. Placing dynamic content in caches with small population. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

[45] Yuanyuan Li, Tareq Si Salem, Giovanni Neglia, and Stratis Ioannidis. 2021. Online Caching Networks with Adversarial Guarantees. *Proc. ACM Meas. Anal. Comput. Syst.* 5, 3, Article 35 (dec 2021), 39 pages.

[46] Luofeng Liao, Yuan Gao, and Christian Kroer. 2022. Nonstationary Dual Averaging and Online Fair Allocation. *ArXiv e-prints* (Feb. 2022).

[47] Luyang Liu, Hongyu Li, and Marco Gruteser. 2019. Edge Assisted Real-time Object Detection for Mobile Augmented Reality. In *MobiCom '19: The 25th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 1–16.

[48] Yuezhou Liu, Yuanyuan Li, Qian Ma, Stratis Ioannidis, and Edmund Yeh. 2020. Fair Caching Networks. *Performance Evaluation* 143 (2020), 102138.

[49] Andrea Lodi, Philippe Olivier, Gilles Pesant, and Sriram Sankaranarayanan. 2021. Fairness over time in dynamic resource allocation with an application in Healthcare. *arXiv preprint arXiv:2101.03716* (2021).

[50] Shie Mannor, John N Tsitsiklis, and Jia Yuan Yu. 2009. Online Learning with Sample Path Constraints. *Journal of Machine Learning Research* 10, 3 (2009).

[51] H Brendan McMahan. 2017. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research* 18, 1 (2017), 3117–3166.

[52] Naram Mhaisen, George Iosifidis, and Douglas Leith. 2022. Online Caching with no Regret: Optimistic Learning via Recommendations. *arXiv preprint arXiv:2204.09345* (2022).

[53] Jeonghoon Mo and Jean Walrand. 2000. Fair End-to-End Window-Based Congestion Control. *IEEE/ACM Transactions on networking* 8, 5 (2000), 556–567.

[54] Dritan Nace and Michal Pioro. 2008. Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial. *IEEE Communications Surveys & Tutorials* 10, 4 (2008), 5–17.

[55] John F. Nash. 1950. The Bargaining Problem. *Econometrica* 18, 2 (1950), 155–162.

[56] Debjit Paria and Abhishek Sinha. 2021. LeadCache: Regret-Optimal Caching in Networks. *Advances in Neural Information Processing Systems* 34 (2021), 4435–4447.

[57] Georgios S Paschos, Apostolos Destounis, Luigi Vigneri, and George Iosifidis. 2019. Learning to cache with no regrets. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 235–243.
[58] Bozidar Radunovic and Jean-Yves Le Boudec. 2007. A unified framework for max-min and min-max fairness with applications. *IEEE/ACM Transactions on networking* 15, 5 (2007), 1073–1083.
[59] Alexander Rakhlin and Karthik Sridharan. 2013. Online learning with predictable sequences. In *Conference on Learning Theory*. PMLR, 993–1019.
[60] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. 2011. Online Learning: Beyond Regret. In *Proceedings of the 24th Annual Conference on Learning Theory (Proceedings of Machine Learning Research, Vol. 19)*, Sham M. Kakade and Ulrike von Luxburg (Eds.). PMLR, Budapest, Hungary, 559–594.
[61] Resource Allocation and Cross Layer Control in Wireless Networks. 2006. L. Georgiadis, M. J. Neely, and L. Tassiulas. *Found. Trends Netw.* 1, 1 (2006), 1–143.
[62] Adrian Rivera, He Wang, and Huan Xu. 2018. The Online Saddle Point Problem and Online Convex Optimization with Knapsacks. *preprint arXiv:1806.08301* (2018).
[63] Shai Shalev-Shwartz. 2012. Online Learning and Online Convex Optimization. *Foundations and Trends in Machine Learning* 4, 2 (Feb. 2012), 107–194.
[64] Shai Shalev-Shwartz and Yoram Singer. 2007. *Online Learning: Theory, Algorithms, and Applications*. Ph.D. Dissertation. Hebrew University.
[65] Karthikeyan Shanmugam, Negin Golrezaei, Alexandros G. Dimakis, Andreas F. Molisch, and Giuseppe Caire. 2013. FemtoCaching: Wireless Content Delivery Through Distributed Caching Helpers. *IEEE Transactions on Information Theory* 59, 12 (2013), 8402–8413.
[66] Tareq Si Salem, Giovanni Neglia, and Damiano Carra. 2021. AÇAI: Ascent Similarity Caching with Approximate Indexes. In *2021 33rd International Teletraffic Congress (ITC-33)*. IEEE, 1–9.
[67] Tareq Si Salem, Giovanni Neglia, and Stratis Ioannidis. 2021. No-Regret Caching via Online Mirror Descent. In *ICC 2021-IEEE International Conference on Communications*. IEEE, 1–6.
[68] Tareq Si Salem, Giovanni Neglia, and Stratis Ioannidis. 2021. No-Regret Caching via Online Mirror Descent. *preprint arXiv:2101.12588* (2021). arXiv:2101.12588 [cs.LG]
[69] Sean R. Sinclair, Siddhartha Banerjee, and Christina Lee Yu. 2022. Sequential Fair Allocation: Achieving the Optimal Envy-Efficiency Tradeoff Curve. *SIGMETRICS Perform. Eval. Rev.* 50, 1 (jun 2022), 95–96.
[70] Sean R Sinclair, Gauri Jain, Siddhartha Banerjee, and Christina Lee Yu. 2020. Sequential fair allocation of limited resources under stochastic demands. *arXiv preprint arXiv:2011.14382* (2020).
[71] T. Bonald, and J. W. Roberts. 2015. Multi-Resource Fairness: Objectives, Algorithms and Performance. In *ACM Sigmetrics*.
[72] Mohammad Sadegh Talebi and Alexandre Proutiere. 2018. Learning Proportionally Fair Allocations with Low Regret. *SIGMETRICS Performance Evaluation Review* 46, 1 (jun 2018), 50–52.
[73] Stefano Traverso, Mohamed Ahmed, Michele Garetto, Paolo Giaccone, Emilio Leonardi, and Saverio Niccolini. 2013. Temporal locality in today's content caching: Why it matters and how to model it. *ACM SIGCOMM Computer Communication Review* 43, 5 (2013), 5–12.
[74] W. Wang, B. Li, and B. Liang. 2014. Dominant Resource Fairness in Cloud Computing Systems with Heterogeneous Servers. In *IEEE INFOCOM*.
[75] Liang Wang, Gareth Tyson, Jussi Kangasharju, and Jon Crowcroft. 2017. Milking the Cache Cow With Fairness in Mind. *IEEE/ACM Transactions on Networking* 25, 5 (2017), 2686–2700.
[76] Weiran Wang and Canyi Lu. 2015. Projection onto the Capped Simplex. *preprint arXiv:1503.01002* (2015).
[77] X. Lin, N. B. Shroff, and R. Srikant. 2006. A Tutorial on Cross-Layer Optimization in Wireless Networks. *IEEE J. Sel. Areas Commun.* 24, 8 (2006), 1452–1463.
[78] David Zeng and Alexandros Psomas. 2020. Fairness-efficiency tradeoffs in dynamic fair division. In *Proceedings of the 21st ACM Conference on Economics and Computation*. 911–912.
[79] Jiao Zhang, F. Richard Yu, Shuo Wang, Tao Huang, Zengyi Liu, and Yunjie Liu. 2018. Load Balancing in Data Center Networks: A Survey. *IEEE Communications Surveys & Tutorials* 20, 3 (2018), 2324–2352.
[80] Yu-Hang Zhou, Chen Liang, Nan Li, Cheng Yang, Shenghuo Zhu, and Rong Jin. 2019. Robust online matching with user arrival distribution drift. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 459–466.

A Technical Lemmas and Definitions

A.1 Convex Conjugate

Definition 3. *Let* $F : \mathcal{U} \subset \mathbb{R}^{\mathcal{I}} \to \mathbb{R} \cup \{-\infty, +\infty\}$ *be a function. Define* $F^\star : \mathbb{R}^{\mathcal{I}} \to \mathbb{R} \cup \{-\infty, +\infty\}$ *by*

$$F^\star(\boldsymbol{\theta}) = \sup_{\boldsymbol{u} \in \mathcal{U}} \{\boldsymbol{u} \cdot \boldsymbol{\theta} - F(\boldsymbol{u})\}, \tag{28}$$

for $\boldsymbol{\theta} \in \mathbb{R}^{\mathcal{I}}$. *This is the* convex conjugate *of* F.

A.2 Convex Conjugate of α-Fairness Function

Lemma 1. *Let* $\mathcal{U} = [u_{\star,\min}, u_{\star,\max}]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{>0}$, $\Theta = \left[-1/u^\alpha_{\star,\min}, -1/u^\alpha_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{<0}$, *and* $F_\alpha : \mathcal{U} \to \mathbb{R}$ *be an* α*-fairness function* (3). *The convex conjugate of* $-F_\alpha$ *is given by*

$$(-F_\alpha)^\star(\boldsymbol{\theta}) = \begin{cases} \sum_{i \in \mathcal{I}} \dfrac{\alpha(-\theta_i)^{1-1/\alpha} - 1}{1-\alpha} & \text{for } \alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}, \\ \sum_{i \in \mathcal{I}} -\log(-\theta_i) - 1 & \text{for } \alpha = 1, \end{cases} \tag{29}$$

where $\boldsymbol{\theta} \in \Theta$.

Proof. The convex conjugate of $-f_\alpha(u) \triangleq -\frac{u^{1-\alpha}-1}{1-\alpha}$ for $u \in [u_{\star,\min}, u_{\star,\max}]$ and $\alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}$ is given by

$$(-f_\alpha)^\star(\theta) = \max_{u \in [u_{\star,\min}, u_{\star,\max}]} \left\{ u\theta + \frac{u^{1-\alpha} - 1}{1-\alpha} \right\}. \tag{30}$$

We evaluate the derivative to characterize the maxima of r.h.s. term in the above equation

$$\frac{\partial}{\partial u}\left(u\theta + \frac{u^{1-\alpha}-1}{1-\alpha}\right) = \theta + \frac{1}{u^\alpha}. \tag{31}$$

The function $\theta + \frac{1}{u^\alpha}$ is a decreasing function in u; thus $\theta + \frac{1}{u^\alpha} \geq 0$ when $u \leq \left(-\frac{1}{\theta}\right)^{\frac{1}{\alpha}}$, and $\theta + \frac{1}{u^\alpha} < 0$ otherwise. The maximum is achieved at $u = \left(-\frac{1}{\theta}\right)^{\frac{1}{\alpha}}$. It holds through Eq. (30)

$$(-f_\alpha)^\star(\theta) = \frac{\alpha(-\theta)^{1-1/\alpha} - 1}{1-\alpha} \qquad \text{for} \qquad \theta \in \left[-1/u^\alpha_{\star,\min}, -1/u^\alpha_{\star,\max}\right]. \tag{32}$$

Moreover, it is easy to check that the same argument holds for $f_1(u) = \log(u)$ and we have

$$(-f_1)^\star(\theta) = -1 - \log(-\theta) \qquad \text{for} \qquad \theta \in \left[-1/u_{\star,\min}, -1/u_{\star,\max}\right]. \tag{33}$$

The convex conjugate of $-F_\alpha(\boldsymbol{u}) = \sum_{i \in \mathcal{I}} f_\alpha(u_i)$ for $\boldsymbol{u} \in \mathcal{U}$, using Eq. (32) and Eq. (33), is given by

$$(-F_\alpha)^\star(\boldsymbol{\theta}) = \sum_{i \in \mathcal{I}} (-f_\alpha)^\star(\theta_i) = \begin{cases} \sum_{i \in \mathcal{I}} \dfrac{\alpha(-\theta_i)^{1-1/\alpha} - 1}{1-\alpha} & \text{for } \alpha \in \mathbb{R}_{\geq 0} \setminus \{1\}, \\ \sum_{i \in \mathcal{I}} -\log(-\theta_i) - 1 & \text{for } \alpha = 1, \end{cases} \tag{34}$$

for $\boldsymbol{\theta} \in \Theta$, because $F_\alpha(\boldsymbol{u})$ is separable in $\boldsymbol{u} \in \mathcal{U}$.

■

A.3 Convex Biconjugate of α-Fairness Functions

The following Lemma provides a stronger condition on $\boldsymbol{\theta}$ compared to [1, Lemma 2.2], i.e., we restrict $\boldsymbol{\theta} \in \Theta$ instead of $\|\boldsymbol{\theta}\|_\star \leq L$ where $L \geq \|\nabla_{\boldsymbol{u}} F_\alpha(\boldsymbol{u})\|_\star$ for all $\boldsymbol{u} \in \mathcal{U}$ and $\|\cdot\|_\star$ is the dual norm of $\|\cdot\|$.

Lemma 2. *Let* $\mathcal{U} = \left[u_{\star,\min}, u_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{>0}$, $\Theta = \left[-1/u^{\alpha}_{\star,\min}, -1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{<0}$, *and* $F_\alpha : \mathcal{U} \to \mathbb{R}$ *be an* α*-fairness function* (3). *The function* F_α *can be always be recovered from the convex conjugate* $(-F_\alpha)^\star$, *i.e.,*

$$F_\alpha(\boldsymbol{u}) = \min_{\boldsymbol{\theta}\in\Theta} \left\{(-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta}\cdot\boldsymbol{u}\right\}, \tag{35}$$

for $\boldsymbol{u} \in \mathcal{U}$.

Proof. This proof follows the same lines of the proof of [1, Lemma 2.2]. Since $\boldsymbol{u} \in \mathcal{U}$, therefore the gradient of F_α at point $\boldsymbol{u}$ is given as $\nabla_{\boldsymbol{u}} F_\alpha(\boldsymbol{u}) = \left[1/u_i^\alpha\right]_{i\in\mathcal{I}} \in -\Theta = \left[1/u^{\alpha}_{\star,\min}, 1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}}$. Moreover, it holds

$$\begin{aligned}
\min_{\boldsymbol{\theta}\in\Theta} \left\{(-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta}\cdot\boldsymbol{u}\right\} &= \min_{\boldsymbol{\theta}\in\Theta}\left\{\max_{\boldsymbol{u}'\in\mathcal{U}}\left\{\boldsymbol{\theta}\cdot\boldsymbol{u}' + F_\alpha(\boldsymbol{u})\right\} - \boldsymbol{\theta}\cdot\boldsymbol{u}\right\} && (36)\\
&= \max_{\boldsymbol{u}'\in\mathcal{U}}\min_{\boldsymbol{\theta}\in\Theta}\left\{\boldsymbol{\theta}\cdot\boldsymbol{u}' + F_\alpha(\boldsymbol{u}') - \boldsymbol{\theta}\cdot\boldsymbol{u}\right\}. \qquad \text{Minmax theorem} && (37)
\end{aligned}$$

We take

$$\begin{aligned}
\min_{\boldsymbol{\theta}\in\Theta}\left\{\boldsymbol{\theta}\cdot\boldsymbol{u}' + F_\alpha(\boldsymbol{u}') - \boldsymbol{\theta}\cdot\boldsymbol{u}\right\} &= \min_{\boldsymbol{\theta}\in\Theta}\left\{F_\alpha(\boldsymbol{u}') + \boldsymbol{\theta}\cdot(\boldsymbol{u}'-\boldsymbol{u})\right\}\\
&\leq F_\alpha(\boldsymbol{u}') - \nabla F_\alpha(\boldsymbol{u})\cdot(\boldsymbol{u}'-\boldsymbol{u}) && \text{Because } -\nabla F_\alpha(\boldsymbol{u}) \in \Theta\\
&\leq F_\alpha(\boldsymbol{u}). && \text{Use concavity of } F_\alpha
\end{aligned}$$

The equality is achieved when $\boldsymbol{u} = \boldsymbol{u}'$ and the maximum value in (37) is attained for this value. We conclude the proof. ∎

A.4 Online Gradient Descent (OGD) with Self-Confident Learning Rates

Lemma 3 provides the regret guarantee of OGD oblivious to the time horizon T and bound on subgradients' norm for any $t \in \mathcal{T}$. This adopts the idea of [7] which denominate such learning schemes as self-confident.

Lemma 3. *Consider a convex set* $\mathcal{X}$, *a sequence of* σ*-strongly convex functions* $f_t : \mathcal{X} \to \mathbb{R}$ *with subgradient* $\boldsymbol{g}_t \in \partial f_t(\boldsymbol{x}_t)$ *at* $\boldsymbol{x}_t$, *and OGD update rule* $\boldsymbol{x}_{t+1} = \Pi_{\mathcal{X}}\left(\boldsymbol{x}_t - \eta_t\boldsymbol{g}_t\right) = \arg\min_{\boldsymbol{x}\in\mathcal{X}}\left\|\boldsymbol{x} - \left(\boldsymbol{x}_t - \eta_t\boldsymbol{g}_t\right)\right\|_2$ *initialized at* $\boldsymbol{x}_1 \in \mathcal{X}$. *Let* $\operatorname{diam}(\mathcal{X}) \triangleq \max\left\{\|\boldsymbol{x}-\boldsymbol{x}'\|_2 : \boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}\right\}$. *Selecting the learning rates as* $\boldsymbol{\eta} : \mathcal{T} \to \mathbb{R}$ *such that* $\eta_t \leq \eta_{t-1}$ *for all* $t > 1$ *gives the following regret guarantee against a fixed decision* $\boldsymbol{x} \in \mathcal{X}$*:*

$$\sum_{t\in\mathcal{T}} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \leq \operatorname{diam}^2(\mathcal{X})\sum_{t=1}^{T}\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} - \sigma\right) + \sum_{t=1}^{T}\eta_t\left\|\boldsymbol{g}_t\right\|_2^2. \tag{38}$$

- *When* $\sigma > 0$, *selecting the learning rate schedule* $\eta_t = \frac{1}{\sigma t}$ *for* $t \in \mathcal{T}$ *gives*

$$\sum_{t\in\mathcal{T}} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \leq \sum_{t=1}^{T}\frac{\left\|\boldsymbol{g}_t\right\|_2^2}{t\sigma} = \mathcal{O}\left(\log(T)\right). \tag{39}$$

- *When $\sigma = 0$, selecting the learning rate schedule $\eta_t = \frac{\operatorname{diam}(\mathcal{X})}{\sqrt{\sum_{s=1}^{t} \|\boldsymbol{g}_s\|_2^2}}$ for $t \in \mathcal{T}$ gives*

$$\sum_{t\in\mathcal{T}} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \leq 1.5 \operatorname{diam}(\mathcal{X}) \sqrt{\sum_{t\in\mathcal{T}} \|\boldsymbol{g}_s\|_2^2} = \mathcal{O}\left(\sqrt{T}\right). \tag{40}$$

Proof. This proof follows the same lines of the proof of [34]. We do not assume a bound on the gradients is known beforehand and the time horizon T. Take a fixed $\boldsymbol{x} \in \mathcal{X}$. Applying the definition of σ-strong convexity to the pair of points $\boldsymbol{x}_t$ and $\boldsymbol{x}$, we have

$$2\left(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x})\right) \leq 2\boldsymbol{g}_t \cdot (\boldsymbol{x}_t - \boldsymbol{x}) - \sigma \|\boldsymbol{x}_t - \boldsymbol{x}\|_2^2. \tag{41}$$

Pythagorean theorem implies

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|_2^2 = \left\|\Pi_{\mathcal{X}}\left(\boldsymbol{x}_t - \eta_t \boldsymbol{g}_t\right) - \boldsymbol{x}\right\|_2^2 \leq \|\boldsymbol{x}_t - \eta_t - \boldsymbol{x}\|_2^2, \tag{42}$$

Expanding the r.h.s. term gives

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|_2^2 \leq \|\boldsymbol{x}_t - \boldsymbol{x}\|_2^2 + \eta_t^2 \left\|\boldsymbol{g}_t\right\|_2^2 - 2\eta_t \boldsymbol{g}_t \cdot (\boldsymbol{x}_t - \boldsymbol{x}), \tag{43}$$

$$2\boldsymbol{g}_t \cdot (\boldsymbol{x}_t - \boldsymbol{x}) \leq \frac{\|\boldsymbol{x}_t - \boldsymbol{x}\|_2^2 - \|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|_2^2}{\eta_t} + \eta_t \left\|\boldsymbol{g}_t\right\|_2^2. \tag{44}$$

Combine Eq. (41) and Eq. (44) and for $t = 1$ to $t = T$:

$$\begin{aligned}
2\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) &\leq \sum_{t=1}^{T} \frac{\|\boldsymbol{x}_t - \boldsymbol{x}\|_2^2 (1 - \sigma\eta_t) - \|\boldsymbol{x}_{t+1} - \boldsymbol{x}\|_2^2}{\eta_t} + \sum_{t=1}^{T} \eta_t \left\|\boldsymbol{g}_t\right\|_2^2 \\
&\leq \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{x}\|_2^2 \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} - \sigma\right) + \sum_{t=1}^{T} \eta_t \left\|\boldsymbol{g}_t\right\|_2^2 && \frac{1}{\eta_0} \triangleq 0 \\
&\leq \operatorname{diam}^2(\mathcal{X}) \left(\frac{1}{\eta_T} - \sigma T\right) + \sum_{t=1}^{T} \eta_t \left\|\boldsymbol{g}_t\right\|_2^2. && \text{Telescoping series}
\end{aligned}$$

When $\sigma > 0$ and $\eta_t = \frac{1}{\sigma t}$, from Eq. (45) we have

$$\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \leq 0 + \sum_{t=1}^{T} \frac{\left\|\boldsymbol{g}_t\right\|_2^2}{2\sigma t} \leq \max_{t\in\mathcal{T}} \left\{\left\|\boldsymbol{g}_t\right\|_2^2\right\} \sum_{t=1}^{T} \frac{1}{2\sigma} \leq \frac{\max_{t\in\mathcal{T}}\left\{\left\|\boldsymbol{g}_t\right\|_2^2\right\}}{2\sigma} \mathrm{H}_T = \mathcal{O}\left(\log(T)\right), \tag{45}$$

where H_T is the T-th harmonic number.
When $\sigma = 0$ and $\eta_t = \frac{\operatorname{diam}(\mathcal{X})}{\sqrt{\sum_{s=1}^{t} \|\boldsymbol{g}_s\|_2^2}}$, from Eq. (45) we have

$$\begin{aligned}
\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) &\leq \frac{\operatorname{diam}(\mathcal{X})}{2} \sqrt{\sum_{t=1}^{T} \|\boldsymbol{g}_s\|_2^2} + \frac{\operatorname{diam}(\mathcal{X})}{2} \sum_{t=1}^{T} \frac{\left\|\boldsymbol{g}_t\right\|_2^2}{\sqrt{\sum_{s=1}^{t} \left\|\boldsymbol{g}_s\right\|_2^2}} \qquad (46) \\
&\leq 1.5 \operatorname{diam}(\mathcal{X}) \sqrt{\sum_{t=1}^{T} \|\boldsymbol{g}_s\|_2^2} = \mathcal{O}\left(\sqrt{T}\right). \qquad (47)
\end{aligned}$$

Last inequality is obtained using [7, Lemma 3.5], i.e., $\sum_{t=1}^{T} \frac{|a_t|}{\sum_{s=1}^{t} |a_s|} \leq 2\sqrt{\sum_{t=1}^{T} |a_t|}$. This concludes the proof. ∎

A.5 Saddle-Point Problem Formulation of α-Fairness

Lemma 4. *Let $\mathcal{X}$ be a convex set, $\mathcal{U} = \left[u_{\star,\min}, u_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{>0}$, $u_i : \mathcal{X} \to \mathcal{U}$ be a concave function for every $i \in \mathcal{I}$, $\Theta = \left[-1/u^{\alpha}_{\star,\min}, -1/u^{\alpha}_{\star,\max}\right]^{\mathcal{I}} \subset \mathbb{R}^{\mathcal{I}}_{<0}$, and $\Psi_\alpha : \Theta \times \mathcal{X} \to \mathbb{R}$ be a function given by*

$$\Psi_\alpha(\boldsymbol{\theta}, \boldsymbol{x}) \triangleq (-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \boldsymbol{u}(\boldsymbol{x}). \tag{48}$$

The following holds:

- *The solution of the saddle-point problem formed by Ψ_α is a maximizer of the α-fairness function*

$$\max_{\boldsymbol{x}\in\mathcal{X}} \min_{\boldsymbol{\theta}\in\Theta} \Psi_\alpha(\boldsymbol{\theta}, \boldsymbol{x}) = \max_{\boldsymbol{x}\in\mathcal{X}} F_\alpha(\boldsymbol{u}(\boldsymbol{x})). \tag{49}$$

- *The function $\Psi_\alpha : \Theta \times \mathcal{X} \to \mathbb{R}$ is concave over $\mathcal{X}$.*
- *The function $\Psi_\alpha : \Theta \times \mathcal{X} \to \mathbb{R}$ is $\frac{u^{1+1/\alpha}_{\star,\min}}{\alpha}$-strongly convex over Θ w.r.t. $\|\cdot\|_2$ for $\alpha > 0$.*

Proof. Equation (49) is a direct result of Lemma 2. The function Ψ_α is concave over $\mathcal{X}$ because $-\boldsymbol{\theta} \cdot \boldsymbol{u}(\boldsymbol{x})$ is a weighted sum of concave functions with non-negative weights. To prove the strong convexity of Ψ_α w.r.t. $\|\cdot\|_2$, a sufficient condition [64, Lemma 14] is given by $\boldsymbol{\theta'}^T \left(\nabla^2_{\boldsymbol{\theta}} \Psi_\alpha(\boldsymbol{\theta}, \boldsymbol{x})\right) \boldsymbol{\theta'} \geq \sigma \left\|\boldsymbol{\theta'}\right\|_2^2$ for all $\boldsymbol{\theta}, \boldsymbol{\theta'} \in \Theta$, and it holds

$$\boldsymbol{\theta'}^T \left(\nabla^2_{\boldsymbol{\theta}} \Psi_\alpha(\boldsymbol{\theta}, \boldsymbol{x})\right) \boldsymbol{\theta'} = \sum_{i\in\mathcal{I}} \theta_i'^2 \frac{\partial^2}{\partial \theta_i} (-F_\alpha)^\star(\boldsymbol{\theta}) = \sum_{i\in\mathcal{I}} \frac{\theta_i'^2}{\alpha(-\theta_i)^{1+1/\alpha}} \geq \frac{u^{1+1/\alpha}_{\star,\min}}{\alpha} \left\|\boldsymbol{\theta'}\right\|_2^2. \tag{50}$$

This concludes the proof. ■

B Proof of Theorem 1

Proof. Consider two players $\mathcal{I} = \{1, 2\}$, allocation set $\mathcal{X} = [-1, 1]$ for all $t \in \mathcal{T}$. We define $\gamma_T \in [0.4, 1]$, $\psi_T \triangleq \frac{1}{T} \sum_{t=1}^{\gamma_T T} x_t$. We assume w.l.g. $\gamma_T T$ is a natural number. We consider two strategies selected by the adversary:

Strategy 1. The adversary reveals the following utilities:

$$\boldsymbol{u}_t(x) = \begin{cases} (1 + x, 2 - x) & \text{if } t \leq \gamma_T T, \\ (1, 1) & \text{otherwise.} \end{cases} \tag{51}$$

Under the selected utilities, the static optimum attains the following objective

$$\mathrm{OPT}^{\mathrm{S1}} = \max_{x\in\mathcal{X}} \ f_\alpha((1 + x)\gamma_T + (1 - \gamma_T)) + f_\alpha((2 - x)\gamma_T + (1 - \gamma_T)) \tag{52}$$

$$= \max_{x\in\mathcal{X}} \ f_\alpha(1 + \gamma_T x) + f_\alpha(1 + \gamma_T - \gamma_T x). \tag{53}$$

The above objective is concave in x. We can perform a derivative test to characterize its maximum

$$\frac{\partial f_\alpha(1 + \gamma_T x) + F_\alpha(1 + \gamma_T - \gamma_T x)}{\partial x} = \frac{\gamma_T}{(1 + \gamma_T x)^\alpha} - \frac{\gamma_T}{(1 + \gamma_T - \gamma_T x)^\alpha} = 0, \qquad \text{for } x = \frac{1}{2}. \tag{54}$$

Thus, it holds

$$\mathrm{OPT}^{\mathrm{S1}} = 2 f_\alpha(1 + 0.5\gamma_T). \tag{55}$$

The fairness regret denoted by $\mathfrak{R}^{\mathrm{S1}}_T(F_\alpha, \boldsymbol{A})$ under this strategy of a policy $\mathcal{A}$ is given by

$$\mathfrak{R}^{\mathrm{S1}}_T(F_\alpha, \boldsymbol{A}) = \mathrm{OPT}^{\mathrm{S1}} - f_\alpha\left(\frac{1}{T}\left(\sum_{t=1}^{\gamma_T T} 1 + x_t\right) + 1 - \gamma_T\right) - f_\alpha\left(\frac{1}{T}\left(\sum_{t=1}^{\gamma_T T} 2 - x_t\right) + 1 - \gamma_T\right) \tag{56}$$

$$= 2f_\alpha(1 + 0.5\gamma_T) - f_\alpha(1 + \psi_T) - f_\alpha(1 + \gamma_T - \psi_T). \tag{57}$$

Strategy 2. The adversary reveals the following utilities:

$$\boldsymbol{u}_t(x) = \begin{cases} (1 + x, 2 - x) & \text{if } t \leq \gamma_T T, \\ (2, 0) & \text{otherwise.} \end{cases} \tag{58}$$

Under the selected utilities, the static optimum attains the following objective

$$\mathrm{OPT}^{\mathrm{S2}} = \max_{x \in \mathcal{X}} f_\alpha((1 + x)\gamma_T + (1 - \gamma_T)2) + f_\alpha((2 - x)\gamma_T) \tag{59}$$

$$= \max_{x \in \mathcal{X}} f_\alpha(2 - \gamma_T + \gamma_T x) + f_\alpha(2\gamma_T - \gamma_T x). \tag{60}$$

Similar to the previous strategy, we can perform a derivative test to characterize the maximum of the the above objective

$$\frac{\partial f_\alpha(2 - \gamma_T + \gamma_T x) + f_\alpha(2\gamma_T - \gamma_T x)}{\partial x} = \frac{\gamma_T}{(2 - \gamma_T + \gamma_T x)^\alpha} - \frac{\gamma_T}{(2\gamma_T - \gamma_T x)^\alpha} = 0, \qquad \text{for } x = 1.5 - \frac{1}{\gamma_T}.$$

Therefore, it holds

$$\mathrm{OPT}^{\mathrm{S2}} = 2f_\alpha(1 + 0.5\gamma_T). \tag{61}$$

and the fairness regret $\mathfrak{R}^{\mathrm{S2}}_T(F_\alpha, \boldsymbol{A})$ under this strategy is

$$\mathfrak{R}^{\mathrm{S2}}_T(F_\alpha, \boldsymbol{A}) = \mathrm{OPT}^{\mathrm{S2}} - f_\alpha\left(\frac{1}{T}\left(\sum_{t=1}^{\gamma_T T} 1 + x_t\right) + 2 - 2\gamma_T\right) - f_\alpha\left(\frac{1}{T}\left(\sum_{t=1}^{\gamma_T T} 2 - x_t\right)\right) \tag{62}$$

$$= 2f_\alpha(1 + 0.5\gamma_T) - f_\alpha(2 - \gamma_T + \psi_T) - f_\alpha(2\gamma_T - \psi_T). \tag{63}$$

We take the average fairness regret $\frac{1}{2}\left(\mathfrak{R}^{\mathrm{S1}}_T(F_\alpha, \boldsymbol{A}) + \mathfrak{R}^{\mathrm{S2}}_T(F_\alpha, \boldsymbol{A})\right)$ across the two strategies

$$\frac{1}{2}\left(\mathfrak{R}^{\mathrm{S1}}_T(F_\alpha, \boldsymbol{A}) + \mathfrak{R}^{\mathrm{S2}}_T(F_\alpha, \boldsymbol{A})\right) \tag{64}$$

$$= 2f_\alpha(1 + 0.5\gamma_T) - \frac{1}{2}\left(f_\alpha(2 - \gamma_T + \psi_T) + f_\alpha(2\gamma_T - \psi_T) + f_\alpha(1 + \psi_T) + f_\alpha(1 + \gamma_T - \psi_T)\right). \tag{65}$$

The r.h.s. of the above equation is convex in ψ_T, so its minimizer can be characterized through the derivative as follows

$$\frac{\partial f_\alpha(2 - \gamma_T + \psi) + f_\alpha(2\gamma_T - \psi) + f_\alpha(1 + \psi) + f_\alpha(1 + \gamma_T - \psi)}{\partial \psi} \tag{66}$$

$$= \frac{1}{(2 - \gamma_T + \psi)^\alpha} - \frac{1}{(2\gamma_T - \psi)^\alpha} + \frac{1}{(1 + \psi)^\alpha} - \frac{1}{(1 + \gamma_T - \psi)^\alpha} = 0, \qquad \text{for } \psi = \gamma_T - 0.5. \tag{67}$$

We replace $\psi_T = \gamma_T - 0.5$ in Eq. (65) to get

$$\frac{1}{2}\left(\mathfrak{R}^{\mathrm{S1}}_T(F_\alpha, \boldsymbol{A}) + \mathfrak{R}^{\mathrm{S2}}_T(F_\alpha, \boldsymbol{A})\right) \geq 2f_\alpha(1 + 0.5\gamma_T) - \left(f_\alpha(1.5) + f_\alpha(0.5 + \gamma_T)\right). \tag{68}$$

We take the derivative of the lower bound

$$\frac{\partial 2f_\alpha(1 + 0.5\gamma_T) - (f_\alpha(1.5) + f_\alpha(0.5 + \gamma_T))}{\partial \gamma_T} = \frac{(0.5 + \gamma)^\alpha - (1 + 0.5\gamma)^\alpha}{(0.5 + \gamma)^\alpha\,(1 + 0.5\gamma)^\alpha}. \tag{69}$$

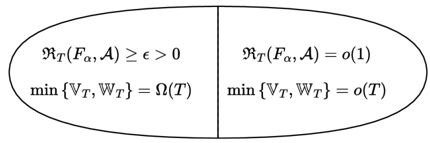


Fig. 12. Assumption (A5) and fairness regret (8) under scenarios 1 and 2.

Note that the sign of the derivative is determined by the numerator $(0.5+\gamma)^\alpha - (1+0.5\gamma)^\alpha$. It holds $(0.5+\gamma)^\alpha - (1+0.5\gamma)^\alpha < 0$ for $\gamma_T < 1$, otherwise $(0.5+\gamma)^\alpha - (1+0.5\gamma)^\alpha = 0$. Hence, the lower bound in Eq. (68) is strictly decreasing for $\gamma_T < 1$, and it holds for $\gamma_T \leq 1-\epsilon$ for $\epsilon > 0$

$$\frac{1}{2}\left(\mathfrak{R}_T^{\mathrm{S1}}(F_\alpha, \boldsymbol{A}) + \mathfrak{R}_T^{\mathrm{S2}}(F_\alpha, \boldsymbol{A})\right) \geq 2f_\alpha(1.5-0.5\epsilon) - (f_\alpha(1.5) + f_\alpha(1.5+0.5\epsilon)) > 0. \tag{70}$$

In other words, the fairness regret guarantee is not attainable[7] for values of $\gamma_T \leq 1-\epsilon$ for any T and $\epsilon > 0$. We can also verify that (A5) is violated when $\gamma_T \leq 1-\epsilon$ for any T and $\epsilon > 0$. Note that γ_T is defined to be in the set $[0.4, 1]$.

Under strategy 1 we have $x_\star = \frac{1}{2}$ and it holds

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(x_\star) = (1+0.5\gamma_T, 1+0.5\gamma_T), \text{ and } \boldsymbol{u}_t(x_\star) = \begin{cases}(1.5, 1.5) & \text{if } t \leq \gamma_T T,\\ (1,1) & \text{otherwise.}\end{cases} \tag{71}$$

Then, it holds

$$\mathbb{V}_{\mathcal{T}} \geq 2(1-\gamma_t)\gamma_T T \geq 2\epsilon\gamma_T T \geq 0.8\epsilon T = \Omega(T). \tag{72}$$

Moreover, it can easily be checked that $\mathbb{W}_{\mathcal{T}} = \Omega(T)$ because there is no decomposition $\{1,2,\ldots,T\} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \cdots \cup \mathcal{T}_K$ where $\max\{\mathcal{T}_k : k \in [K]\} = o\left(T^{\frac{1}{2}}\right)$ under which $\sum_{k=1}^{K}\sum_{i\in\mathcal{I}}\left|\sum_{t\in\mathcal{T}_k}\delta_{t,i}(\boldsymbol{x}_\star)\right| = o(T)$.

To conclude, when $\gamma_T = 1-o(1)$, we have $\min\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\} \leq \mathbb{V}_{\mathcal{T}} = o(T)$; thus, Assumption (A5) only holds when $\gamma_T = 1-o(1)$ for which the vanishing fairness regret guarantee is attainable. Figure 12 provides a summary of the connection between the fairness regret under scenarios 1 and 2 and Assumption (A5). ∎

C Proof of Theorem 2

Proof. Note that $\Psi_{\alpha,t} : \Theta \times \mathcal{X} \to \mathbb{R}$ is the function given by

$$\Psi_{\alpha,t}(\boldsymbol{\theta}, \boldsymbol{x}) = (-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta}\cdot\boldsymbol{u}_t(\boldsymbol{x}), \tag{73}$$

where $F_\alpha : \mathcal{U} \to \mathbb{R}$ is an α-fairness function (3). From Lemma 3, OGD operating over the set Θ under the $\frac{u_{\star,\min}^{1+1/\alpha}}{\alpha}$-strongly convex (Lemma 4) cost functions $\Psi_{\alpha,t}(\underline{\boldsymbol{\theta}_t}, \boldsymbol{x}_t)$ has the following regret guarantee against any fixed $\boldsymbol{\theta} \in \Theta$

$$\frac{1}{T}\sum_{t=1}^{T}\Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) - \frac{1}{T}\sum_{t=1}^{T}\Psi_{\alpha,t}(\boldsymbol{\theta}, \boldsymbol{x}_t) \leq \frac{1}{T}\cdot\underbrace{\frac{1}{2}\sum_{t=1}^{T}\frac{\alpha}{u_{\star,\min}^{1+1/\alpha}t}\left\|\boldsymbol{g}_{\Theta,t}\right\|_2^2}_{\mathfrak{R}_{T,\Theta}}, \tag{74}$$

[7]Note that the fairness regret must vanish for any adversarial choice of sequence of utilities.

From Lemma 2, it holds

$$\min_{\boldsymbol{\theta} \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}, \boldsymbol{x}_t) = F_\alpha \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_t) \right). \tag{75}$$

Combine Eq. (74) and Eq. (75) to obtain the lower bound

$$F_\alpha \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_t) \right) + \frac{\mathfrak{R}_{T,\Theta}}{T} \geq \frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_t). \tag{76}$$

From Lemma 3, OGD operating over the set $\mathcal{X}$ under the reward functions $\Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x})$ has the following regret guarantee for any fixed $\boldsymbol{x}_\star \in \mathcal{X}$:

$$\frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_\star) - \frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) \leq \frac{1}{T} \cdot \underbrace{1.5 \operatorname{diam}(\mathcal{X}) \sqrt{\sum_{t \in \mathcal{T}} \left\| \boldsymbol{g}_{\mathcal{X},t} \right\|_2^2}}_{\mathfrak{R}_{T,\mathcal{X}}}, \tag{77}$$

Hence, we have the following

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_t) + \frac{\mathfrak{R}_{T,\mathcal{X}}}{T} \geq \frac{1}{T} \sum_{t=1}^{T} \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_\star) \\
&= \frac{1}{T} \sum_{t=1}^{T} F^\star(\boldsymbol{\theta}_t) - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\theta}_t \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star) && \text{Replace } \Psi_{\alpha,t}(\boldsymbol{\theta}_t, \boldsymbol{x}_\star) \text{ using Eq. (73)} \\
&\geq F^\star \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\theta}_t \right) - \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\theta}_t \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star) && \text{Jensen's inequality \& convexity of } F^\star \\
&\geq F^\star \left(\bar{\boldsymbol{\theta}} \right) - \bar{\boldsymbol{\theta}} \cdot \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_\star) \right) - \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star) \\
&\geq \min_{\boldsymbol{\theta} \in \Theta} \left\{ F^\star(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_\star) \right) \right\} - \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}) \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star) \\
&= F_\alpha \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_\star) \right) - \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}} \right) \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star).
\end{aligned} \tag{78}$$

We combine the above equation and Eq. (76) to obtain

$$\begin{aligned}
F_\alpha \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_\star) \right) - F_\alpha \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_t) \right) &\leq \frac{\mathfrak{R}_{T,\mathcal{X}}}{T} + \frac{\mathfrak{R}_{T,\Theta}}{T} + \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}} \right) \cdot \boldsymbol{u}_t(\boldsymbol{x}_\star) \\
&= \frac{\mathfrak{R}_{T,\mathcal{X}}}{T} + \frac{\mathfrak{R}_{T,\Theta}}{T} + \underbrace{\frac{1}{T} \sum_{t=1}^{T} \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_t \right) \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star)}_{\Sigma}
\end{aligned} \tag{79}$$

We provide two approaches to bound the r.h.s. term Σ in Eq. (81), and this gives the two conditions in Assumption (A5):

Bound 1. We can bound the r.h.s. term Σ in the above equation as follows

$$\Sigma = \bar{\boldsymbol{\theta}} \cdot \sum_{t=1}^{T} \boldsymbol{\delta}_t(\boldsymbol{x}_\star) - \sum_{t=1}^{T} \boldsymbol{\theta}_t \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star) = -\sum_{t=1}^{T} \boldsymbol{\theta}_t \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star) \tag{80}$$

$$\leq \frac{1}{u_{\star,\min}} \sum_{i \in \mathcal{I}} \sum_{t=1}^{T} \delta_{t,i}(\boldsymbol{x}_\star) \mathbb{1}_{\{\delta_{t,i}(\boldsymbol{x}_\star) \geq 0\}} = O\left(\mathbb{V}_{\mathcal{T}}\right). \tag{81}$$

Bound 2. We alternatively bound Σ as follows

$$\Sigma = \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right) \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star) = \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\min(\mathcal{T}_k)}\right) \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star) + \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left(\boldsymbol{\theta}_{\min(\mathcal{T}_k)} - \boldsymbol{\theta}_t\right) \cdot \boldsymbol{\delta}_t(\boldsymbol{x}_\star)$$

$$\leq \Delta_\alpha \sum_{k=1}^{K} \left\| \sum_{t \in \mathcal{T}_k} \boldsymbol{\delta}_t(\boldsymbol{x}_\star) \right\|_1 + u_{\max} \sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left\| \boldsymbol{\theta}_{\min(\mathcal{T}_k)} - \boldsymbol{\theta}_t \right\|_1, \tag{82}$$

where $\Delta_\alpha = \max\left\{\left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|_\infty : \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta\right\}$. We bound the term $\sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left\|\boldsymbol{\theta}_{\min(\mathcal{T}_k)} - \boldsymbol{\theta}_t\right\|_1$ in the above equation as

$$\sum_{k=1}^{K} \sum_{t \in \mathcal{T}_k} \left\| \boldsymbol{\theta}_{\min(\mathcal{T}_k)} - \boldsymbol{\theta}_t \right\|_1 \leq L_\Theta \sum_{k=1}^{K} \eta_{\Theta, \min(\mathcal{T}_k)} \sum_{t \in \mathcal{T}_k} \left(t - \min\left(\mathcal{T}_k\right)\right) \leq L_\Theta \sum_{k=1}^{K} \eta_{\Theta, \min(\mathcal{T}_k)} \left|\mathcal{T}_k\right|^2 \tag{83}$$

$$= L_\Theta \frac{u_{\star,\min}^{1+\frac{1}{\alpha}}}{\alpha} \sum_{k=1}^{K} \frac{\left|\mathcal{T}_k\right|^2}{\min\left(\mathcal{T}_k\right)}, \tag{84}$$

and replacing this upper-bound in Eq. (82) gives

$$\Sigma \leq \Delta_\alpha \sum_{k=1}^{K} \left\| \sum_{t \in \mathcal{T}_k} \boldsymbol{\delta}_t(\boldsymbol{x}_\star) \right\|_1 + u_{\max} L_\Theta \frac{\alpha}{u_{\star,\min}^{1+\frac{1}{\alpha}}} \sum_{k=1}^{K} \frac{\left|\mathcal{T}_k\right|^2}{\underbrace{\min\left(\mathcal{T}_k\right)}_{\sum_{k'<k} |\mathcal{T}_k| + 1}} = O\left(\mathbb{W}_{\mathcal{T}}\right). \tag{85}$$

We combine Eq. (81), Eq. (85), and Eq. (79) to obtain

$$\mathfrak{R}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) \leq \sup_{\{\boldsymbol{u}_t\}_{t=1}^{T} \in \mathcal{U}^T} \left\{ \frac{1}{T} \left(\mathfrak{R}_{T,\mathcal{X}} + \mathfrak{R}_{T,\Theta}\right) \right\} + O\left(\frac{\min\left\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\right\}}{T}\right) \tag{86}$$

$$\leq \sup_{\{\boldsymbol{u}_t\}_{t=1}^{T} \in \mathcal{U}^T} \left\{ \frac{1}{T} \left(1.5 \operatorname{diam}\left(\mathcal{X}\right) \sqrt{\sum_{t \in \mathcal{T}} \left\|\boldsymbol{g}_{\mathcal{X},t}\right\|_2^2} + \frac{\alpha}{u_{\star,\min}^{1+\frac{1}{\alpha}}} \sum_{t=1}^{T} \frac{\left\|\boldsymbol{g}_{\Theta,t}\right\|_2^2}{t} \right) \right\} + O\left(\frac{\min\left\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\right\}}{T}\right). \tag{87}$$

The following upper bounds hold

$$\left\|\boldsymbol{g}_{\Theta,t}\right\|_2 = \left\| \left(\frac{1}{\left(-\theta_{t,i}\right)^{1/\alpha}} - \left(\boldsymbol{u}_t(\boldsymbol{x}_t)\right) \right)_{i \in \mathcal{I}} \right\|_2 \leq \sqrt{I} \max\left\{ \frac{1}{u_{\star,\min}^{1/\alpha}} - u_{\min}, u_{\max} - \frac{1}{u_{\star,\max}^{1/\alpha}} \right\} = L_\Theta,$$

$$\left\|\boldsymbol{g}_{\mathcal{X},t}\right\|_2 = \left\|\boldsymbol{\theta}_t \cdot \partial_{\boldsymbol{x}} \boldsymbol{u}_t(\boldsymbol{x}_t)\right\|_2 \leq \frac{1}{u_{\star,\min}^{\alpha}} \left\|\partial_{\boldsymbol{x}} \boldsymbol{u}_t(\boldsymbol{x}_t)\right\|_2 \leq \frac{L_{\mathcal{X}}}{u_{\star,\min}^{\alpha}}.$$

Thus, the regret bound in Eq. (87) can be upper bounded as

$$\begin{aligned}
\mathfrak{R}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) &= \frac{1}{T} \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ 1.5 \operatorname{diam}\left(\mathcal{X}\right) \frac{L_{\mathcal{X}}\sqrt{T}}{u_{\star,\min}^{\alpha}} + \frac{\alpha}{u_{\star,\min}^{1+\frac{1}{\alpha}}} \sum_{t=1}^{T} \frac{L_\Theta^2}{t} \right\} + \frac{\min\left\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\right\}}{T} \\
&\leq \frac{1}{T} \sup_{\{\boldsymbol{u}_t\}_{t=1}^T \in \mathcal{U}^T} \left\{ 1.5 \operatorname{diam}\left(\mathcal{X}\right) \frac{L_{\mathcal{X}}\sqrt{T}}{u_{\star,\min}^{\alpha}} + \frac{\alpha}{u_{\star,\min}^{1+\frac{1}{\alpha}}} L_\Theta^2 (\log(T)+1) \right\} + \frac{\min\left\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\right\}}{T} \\
&= \mathcal{O}\left(\frac{1}{\sqrt{T}} + \frac{\min\left\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\right\}}{T} \right).
\end{aligned}$$

This concludes the proof. ∎

D Proof of Theorem 3 (Lower Bound)

Proof. Consider a scenario with a single player $\mathcal{I} = \{1\}$, $\mathcal{X} = \{x \in \mathbb{R}, |x| \leq 1\}$, and the utility selected by an adversary at time slot $t \in \mathcal{T}$ is given by

$$u_t(x) = w_t x + 1, \quad \text{where } w_t \in \{-1, +1\}. \tag{88}$$

The weight w_t is selected in $\{-1, +1\}$ uniformly at random for $t \in \mathcal{T}$. A policy $\mathcal{A}$ selects the sequence of decisions $\{x_t\}_{t=1}^T$ and has the following fairness regret

$$\begin{aligned}
&\mathbb{E}\left[\max_{x\in\mathcal{X}} f_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} u_t(x)\right) - f_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} u_t(x_t)\right)\right] \geq \mathbb{E}\left[\max_{x\in\mathcal{X}} f_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} u_t(x)\right)\right] - \underbrace{f_\alpha\left(\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} u_t(x_t)\right]\right)}_{=0} \\
&= \mathbb{E}\left[f_\alpha\left(\max_{x\in\mathcal{X}} \frac{1}{T}\sum_{t=1}^{T} u_t(x)\right)\right] = \mathbb{E}\left[f_\alpha\left(\frac{1}{T}\left|\sum_{t=1}^{T} w_{t,1}\right| + 1\right)\right] \overset{(a)}{\geq} \mathbb{E}\left[\frac{1}{T}\left|\sum_{t=1}^{T} w_{t,1}\right|\right]\left(\frac{2^{1-\alpha}-1}{1-\alpha}\right) \overset{(b)}{\geq} \frac{\left(\frac{2^{1-\alpha}-1}{1-\alpha}\right)}{\sqrt{2T}} \\
&= \Omega\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Inequality (a) is obtained considering $f_\alpha(x+1)$ is concave in x, $f_\alpha(0+1) = 0$, and $f_\alpha(x+1) \geq f_\alpha(2)x$ for $x \in [0, 1]$. Inequality (b) is obtained through Khintchine inequality. A lower bound on the fairness regret (8) can be established:

$$\mathfrak{R}_T\left(F_\alpha, \boldsymbol{\mathcal{A}}\right) \geq \mathbb{E}\left[\max_{x\in\mathcal{X}} f_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} u_t(x)\right) - f_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} u_t(x_t)\right)\right] = \Omega\left(\frac{1}{\sqrt{T}}\right). \tag{89}$$

This concludes the proof. ∎

E Proof of Corollary 4

Proof. **Expected regret.** When the utilities are i.i.d., we have the following

$$\mathbb{E}\left[\boldsymbol{u}_t(\boldsymbol{x})\right] = \boldsymbol{u}, \forall t \in \mathcal{T}, \tag{90}$$

for some fixed utility $\boldsymbol{u} \in \mathcal{U}$. In the proof Theorem C, in particular, in Eq. (79) it holds

$$F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_\star)\right) - F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}_t(\boldsymbol{x}_t)\right) \leq \frac{\mathfrak{R}_{T,\mathcal{X}}}{T} + \frac{\mathfrak{R}_{T,\Theta}}{T} + \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\theta}_t - \bar{\boldsymbol{\theta}}\right)\cdot \boldsymbol{u}_t(\boldsymbol{x}_\star). \tag{91}$$

Taking the expectation of both sides gives

$$\mathbb{E}\left[F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\boldsymbol{x}_\star)\right)-F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\boldsymbol{x}_t)\right)\right]\leq\mathbb{E}\left[\frac{\mathfrak{R}_{T,\mathcal{X}}}{T}+\frac{\mathfrak{R}_{T,\Theta}}{T}\right]+\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}\right)\cdot\boldsymbol{u}_t(\boldsymbol{x}_\star)\right]. \tag{92}$$

The variables $\boldsymbol{\theta}_t$ and $\boldsymbol{u}_t$ are independent for $t\in\mathcal{T}$, thus we have

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{\theta}_t-\bar{\boldsymbol{\theta}}\right)\cdot\boldsymbol{u}_t(\boldsymbol{x}_\star)\right]=\mathbb{E}\left[\left(\bar{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)\cdot\boldsymbol{u}(\boldsymbol{x}_\star)\right]=0. \tag{93}$$

Through Eq. (88), it holds

$$\mathbb{E}\left[F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\boldsymbol{x}_\star)\right)-F_\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t(\boldsymbol{x}_t)\right)\right]=O\left(\frac{1}{\sqrt{T}}\right). \tag{94}$$

This concludes the first part of the proof.

Almost-sure zero-regret. Let $\Delta=(u_{\max}-u_{\min})$, $\mathcal{T}=\mathcal{T}_1\cup\mathcal{T}_2\cup\cdots\cup\mathcal{T}_K$ where $K=T^{2/3}$ and $|\mathcal{T}_k|=\kappa=T^{1/3}$ for $k\in\{1,2,\ldots,K\}$, and let $\beta\in(0,1/6)$. Employing Hoeffding's inequality we can bound the l.h.s. term in Eq. (9) for $i\in\mathcal{I}$ as

$$\mathbb{P}\left(\left|\sum_{t\in\mathcal{T}_k}\delta_{t,i}(\boldsymbol{x})\right|\leq\Delta T^{1/6+\beta}\right)\geq 1-2\exp\left(\frac{-2T^{1/3+2\beta}}{((T-\kappa)\kappa^2/T^2+\kappa(\kappa/T-1)^2)}\right)=1-2\exp\left(\frac{-2T^{1/3+2\beta}}{(\kappa-\kappa^2/T)}\right) \tag{95}$$

$$=1-2\exp\left(\frac{-2T^{1/3+2\beta}}{\left(T^{1/3}-T^{-1/3}\right)}\right). \tag{96}$$

Hence, it follows

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^{K}\sum_{i\in\mathcal{I}}\left|\sum_{t\in\mathcal{T}_k}\delta_{t,i}(\boldsymbol{x})\right|\leq\Delta T^{5/6+\beta}\right)&\geq\left(1-2\exp\left(\frac{-2T^{1/3+2\beta}}{\left(T^{1/3}-2T^{-1/3}\right)}\right)\right)^{IT^{2/3}}\\&\geq 1-2IT^{2/3}\exp\left(\frac{-2T^{1/3+2\beta}}{\left(T^{1/3}-T^{-1/3}\right)}\right)\qquad\text{Bernoulli's inequality}\\&\geq 1-2IT^{2/3}\exp\left(\frac{-2T^{1/3+2\beta}}{T^{1/3}}\right)\\&\geq 1-2IT^{2/3}\exp(-2T^{2\beta}).\end{aligned}$$

It follows from the above equation paired with Eq. (9)

$$\mathbb{W}_{\mathcal{T}}=O\left(T^{5/6+\beta}+T^{2/3}\right)=O\left(T^{5/6+\beta}\right),\qquad\text{w.p.}\qquad p\geq 1-2IT^{2/3}\exp(-2T^{2\beta}). \tag{97}$$

Thus, for any $\beta\in(0,1/6)$ and $T\to\infty$, it holds

$$\frac{\mathbb{W}_{\mathcal{T}}}{T}\leq 0,\qquad\text{w.p.}\qquad p\geq 1. \tag{98}$$

Note that given that $\mathbb{W}_{\mathcal{T}}\geq 0$ in Eq. (10), it holds $\lim_{T\to\infty}\mathbb{W}_{\mathcal{T}}=0$ w.p. $p=1$. Therefore, it follows from Theorem C for $T\to\infty$

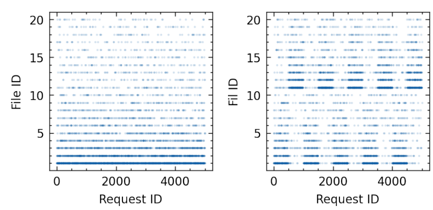
$$\mathfrak{R}_T\left(F_\alpha,\boldsymbol{\mathcal{A}}\right)=O\left(\frac{1}{\sqrt{T}}+\frac{\min\left\{\mathbb{V}_{\mathcal{T}},\mathbb{W}_{\mathcal{T}}\right\}}{T}\right)=O\left(\frac{1}{\sqrt{T}}+\frac{\mathbb{W}_{\mathcal{T}}}{T}\right)\leq 0,\qquad\text{w.p.}\qquad 1. \tag{99}$$

This concludes the proof. ■

F Additional Experimental Details

Table 2. Specification of the network topologies used in experiments.

Topologies	$\lvert\mathcal{C}\rvert$	$\lvert\mathcal{E}\rvert$	k_c	$\lvert\mathcal{Q}_i\rvert$	$\lvert\cup_{f\in\mathcal{F}}\Lambda_f(C)\rvert$	w	Figure
Cycle	3	3	5–5	1	1	1–2	Fig. 5 (a)
Tree-1–Tree-3	13	12	1–5	2–5	1	1–9	Fig. 5 (b)–(d)
Grid	9	12	1–5	2	1	1–7	Fig. 5 (e)
Abilene	12	13	1–5	2	2	1–8	Fig. 5 (f)
GEANT	22	33	1–5	3	2	1–9	Fig. 5 (g)



(a) Stationary (b) Non-Stationary

Fig. 13. Request traces stationary (a) and non-stationary (b) configured with $\sigma = 1.2$, $T = 5000$, $F = 20$, $D = 100$. Each dot indicates a requested file.

G Departing and Arriving Agents

The system model in Section 3.3 supports departing and arriving agents. Consider a population of agents $\mathcal{I}$, the system may only observe a subset of the agents as the *participating* agents, $\mathcal{I}_t \subset \mathcal{I}$ at time t, and the utility of *absent* agents is simply $u_{t,i}(\cdot) = 0$. For example, in the extreme scenario where a single agent $t \in \mathcal{T}$ is participating at a given time slot, the long-term fairness objective (5) falls back to the slot-fairness objective (4), i.e., $F_\alpha\left(\sum_{t\in\mathcal{T}} \boldsymbol{u}_t(\boldsymbol{x}_t)\right) = \sum_{t\in\mathcal{T}} f_\alpha\left(u_{t,t}(\boldsymbol{x}_t)\right)$ where the fairness is ensured across the different agents arriving at different timeslots $t \in \mathcal{T}$. It is easy to verify that even in the case when the set of agents $\mathcal{I}$ is unknown to the controller in advance, one could augment the dual space with an extra dimension each time a new user appears, and the same guarantees hold.

H Time-Complexity of Algorithm 1

Algorithm 1 applied to the virtualized caching system application has a time complexity $O\left(CF^2\right)$, where C is number of caches and F is the number of files in the catalog; the most expensive operation in Algorithm 1 is the projection step in line 8 that corresponds to the Euclidean projection onto a capped simplex, and this can be performed in $O\left(F^2\right)$ steps [76] for each cache state. Despite the high time complexity (F is typically large), in practice solvers (e.g., CVXPY [22]) support *warm-start* that speeds up the projection when the warm-start parameters are close to the ones of obtained by the solution, and since Algorithm 1 is iterative and the cache states do not severely change, typically a lower computational cost is achieved. Moreover, the proposed caching model in Section 6 supports request batching, where a batch includes the requests arriving between two consecutive cache updates. Batching amortizes the computational cost of the different policies, reducing the cost per request by the batch size (R_t at time slot t).

Received August 2022; revised October 2022; accepted November 2022