Enabling Long-term Fairness in Dynamic Resource Allocation

Tareq Si Salem[1] George Iosifidis[2] Giovanni Neglia[3]

[1]Northeastern University (USA) [2]TU Delft (Netherlands) & Amazon (Luxembourg) [3]Inria (France)

Institute for the Wireless Internet of Things at Northeastern

ID23 WIoT Industry Day

Fairness in Resource Allocation

Given a set of agents $\mathcal{I}$ and a vector of utilities $\boldsymbol{u} \in \mathbb{R}^{\mathcal{I}}_{\geq 0}$, the α-fairness criterion is given by

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

It encompasses the utilitarian principle (Bentham-Edgeworth solution) for $\alpha = 0$, the proportional fairness (Nash bargaining solution) for $\alpha = 1$, and the max-min fairness (Kalai–Smorodinsky bargaining solution) for $\alpha = \infty$.

Motivation and Challenge

Today's communication and computing systems require studying fairness in time-varying dynamics, e.g.,

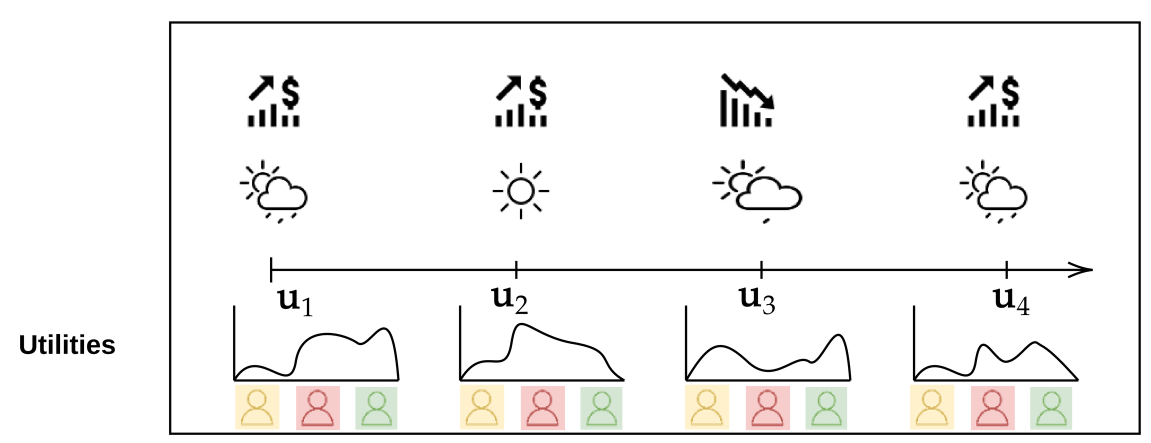
- In **small-cell mobile networks** the user churn is typically very high and unpredictable, thus hindering the fair allocation of spectrum to cells.
- **Caching** files at the edge is non-trivial due to fast-changing patterns of requests.
- **Increasing virtualization** introduces cost and performance volatility.
- **ML and 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.

System Model

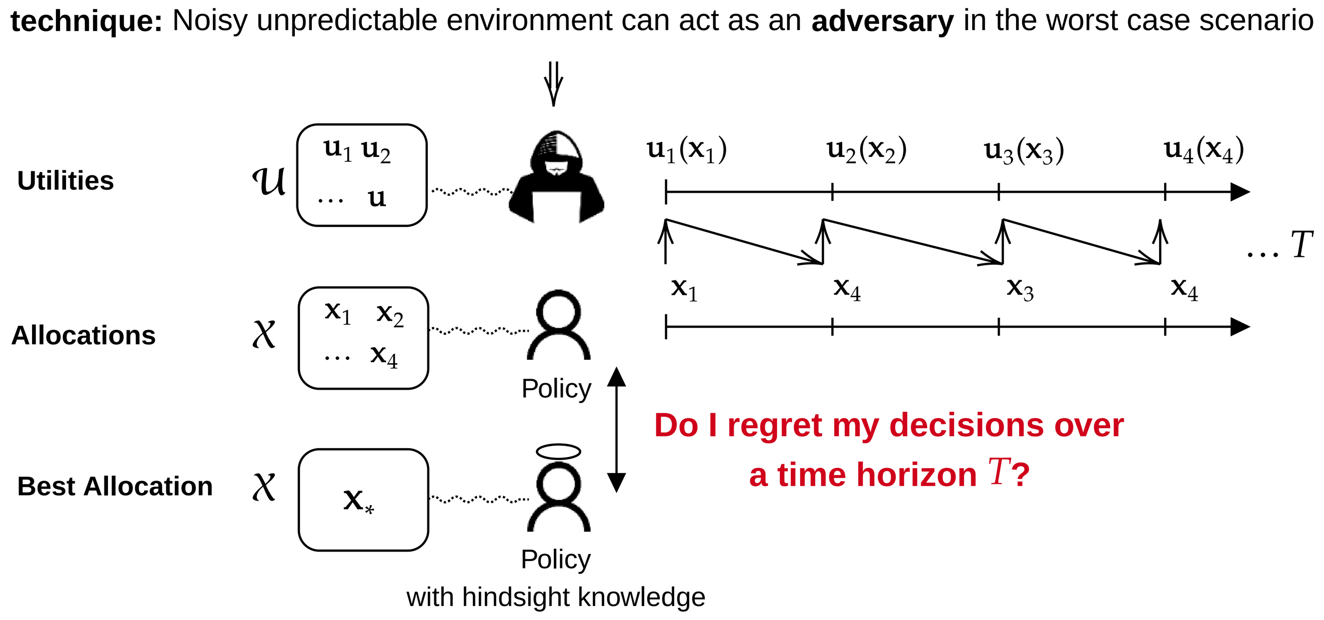
A controller $\mathcal{A}$ selects at each timeslot $t \in \mathbb{N}$ a resource allocation profile $\boldsymbol{x}_t$ from a set of eligible allocations $\mathcal{X}$ based on past agents' utility functions and its previous allocations:

$$\boldsymbol{x}_t = \mathcal{A}_t\left(\{\boldsymbol{x}_s\}_{s=1}^{t-1}, \{\boldsymbol{u}_s\}_{s=1}^{t-1}\right).$$

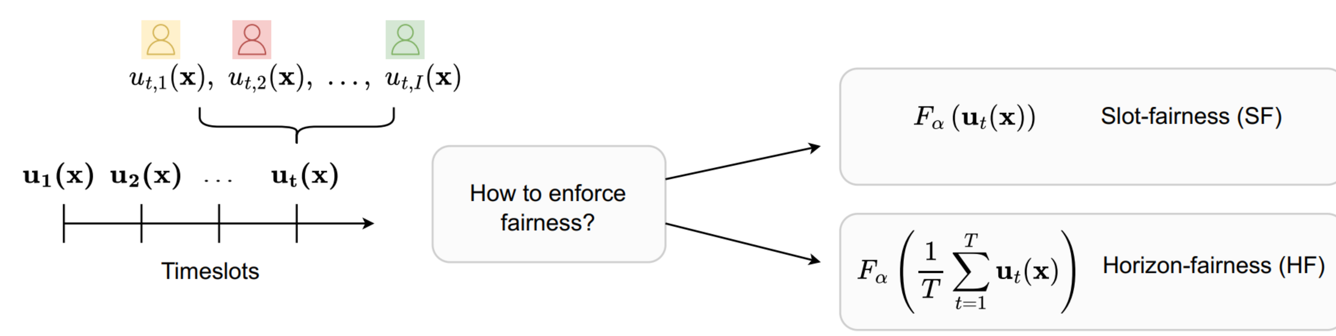
The utilities $\boldsymbol{u}_t$ might change due to unknown, unpredictable, and (possibly) non-stationary perturbations that are revealed to the controller only after it decides its allocation.



Modeling technique: Noisy unpredictable environment can act as an **adversary** in the worst case scenario



- How to define the best allocation?



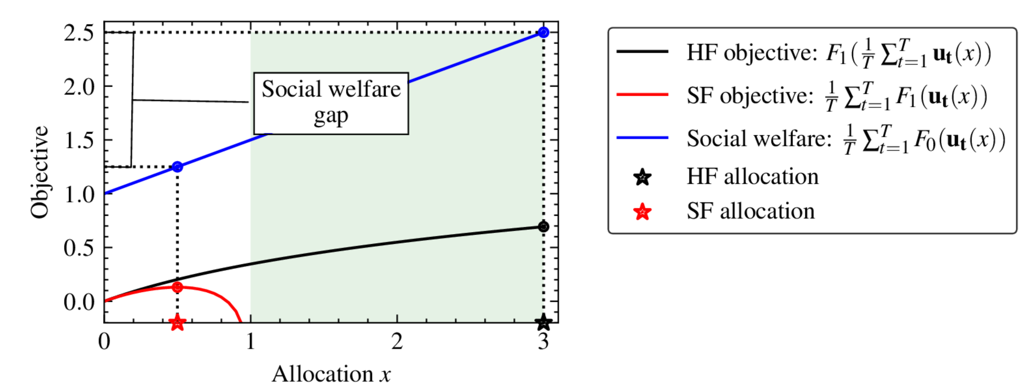


Fig. 3: Example. Consider two agents $\mathcal{I} = \{1, 2\}$, $\alpha = 1$, and a sequence of utilities $\{\boldsymbol{u}_t(x)\}_{t=1}^{T} = \{(1 + x, 1 - x), (1 + x, 1 + x), \dots\}$.

Our focus is on horizon-fairness, which raises novel technical challenges and subsumes slot-fairness as a special case. The performance of a policy $\mathcal{A}$ is evaluated by the *fairness regret*:

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

- We seek a policy $\mathcal{A}$ that guarantees vanishing regret ($\mathfrak{R}_T(F_\alpha, \mathcal{A}) \leq 0$ as $T \to \infty$).

Main Contributions

We first establish an impossibility result.

Impossibility Result

There is no policy $\mathcal{A}$ attaining $\mathfrak{R}_T(F_\alpha, \mathcal{A}) = o(1)$ for $|\mathcal{I}| > 1$ and $\alpha > 0$ under an unrestricted adversary.

We characterize **necessary** and **sufficient** restrictions on the adversary

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

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

The adversary is restricted, so that $\min\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\} = o(T)$. Such a condition captures several practical utility patterns, such as non-stationary corruptions, ergodic, and periodic inputs.

Online Horizon Fair Policy

We formulate an online saddle point problem through the Fenchel convex conjugate

$$\Psi_{t,\alpha}(\boldsymbol{\theta}, \boldsymbol{x}) = (-F_\alpha)^\star(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \boldsymbol{u}_t(\boldsymbol{x}) \qquad \text{for } \boldsymbol{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \Theta,$$

We prove that combining two no-regret policies (gradient ascent in the primal space $\mathcal{X}$ and gradient decent over the dual space Θ) yields the regret guarantee

$$\mathfrak{R}_T(F_\alpha, \mathcal{A}) \leq \mathcal{O}\left(\frac{1}{\sqrt{T}} + \frac{\min\{\mathbb{V}_{\mathcal{T}}, \mathbb{W}_{\mathcal{T}}\}}{T}\right) = o(1).$$

Example Application: Resource Management in Virtualized Caching Systems

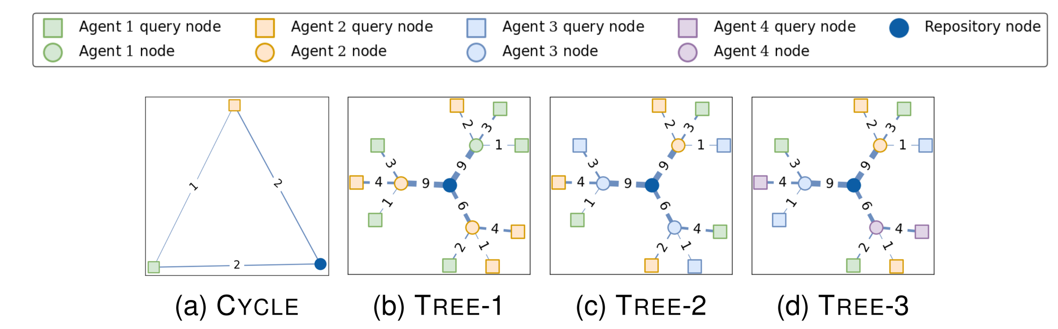


Fig. 4: Network topologies

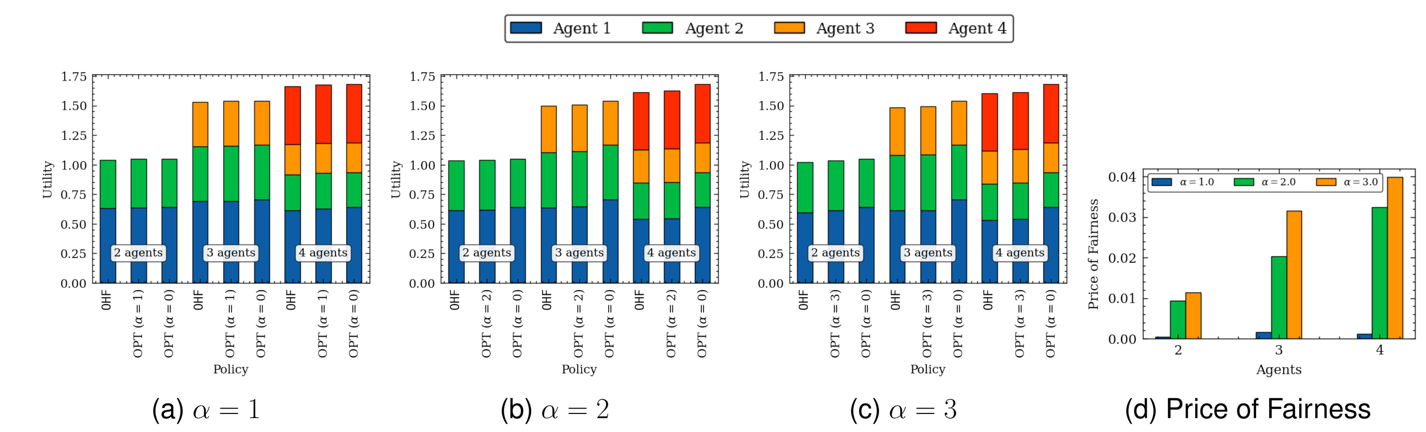


Fig. 5: Price of fairness under different topologies (Tree-1–3) and varying number of participating agents in $\{2, 3, 4\}$.

- We can guarantee horizon-fairness at low price of fairness. The PoF is kept below 4% under the different setups.

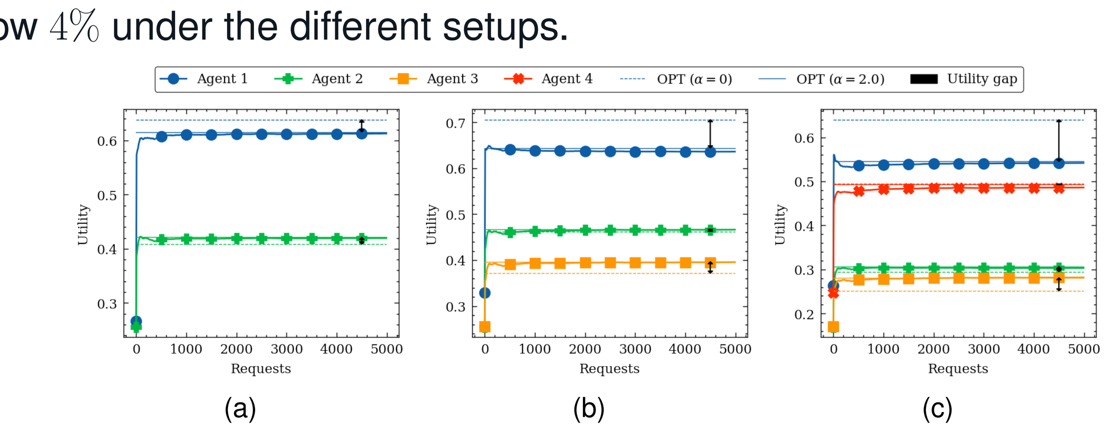


Fig. 6: Convergence speed under different topologies (Tree-1–3)

- The OHF policy is highly adaptive and yields the appropriate allocations in a few iterations.

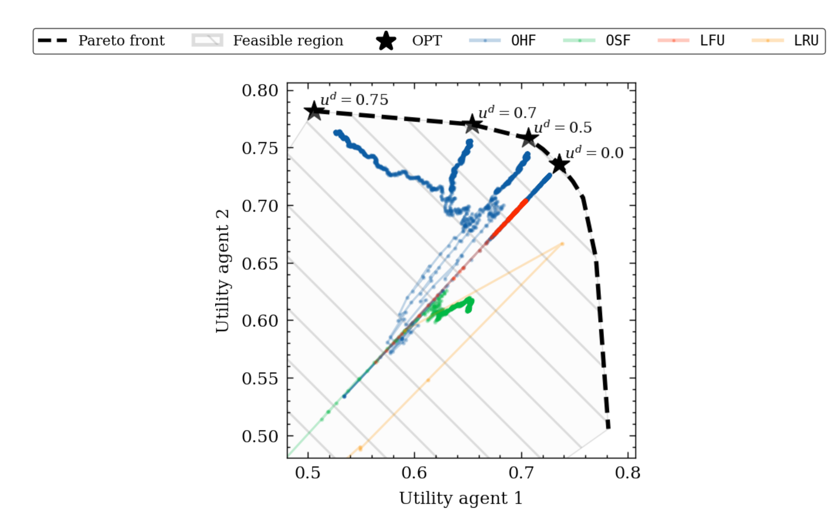


Fig. 7: Nash bargaining ($\alpha = 1$) scenario under different disagreement points and Cycle topology.

- The OHF policy attains solutions that are unfeasible (unreachable) by slot-fairness policies.

Related Publication

This work is published in the proceedings of **ACM SIGMETRICS**, June 19-22, 2023.