Fairness via priority scheduling

Veeraruna Kavitha, N. Hemachandra and Debayan Das
IEOR, IIT Bombay, Mumbai, 400076, India
{vkavitha,nh,debayan}@iitb.ac.in

***Abstract*—In the context of multi-agent resource allocation problems, fairness is a paradigm shift in the recent past. An efficient scheduler always allocates resources to the 'best' agent. Some of the agents, who are most often in 'bad' conditions, are starved and fair schedulers are defined in this context. In this paper, we are interested in the actual gains obtained by the otherwise starved agents, due to fair schedulers. We propose a new notion of fairness, via a constrained optimization, which directly indicates the gains. In general, this constrained optimization is an infinite dimensional problem and the primary contribution of this paper is to reduce it to a tractable finite dimensional zero finding problem. We indicate iterative algorithm(s) which achieves the notion of fairness defined in this paper. We also compare it with some of the existing notions of fairness.**

***Index Terms*—Fairness, Resource allocation, Infinite dimensional convex optimization, Stochastic approximation, Wireless communications**

I. Introduction

In resource allocation problems, agents may have homogeneous (similar) or heterogeneous job requirements. In heterogeneous cases, one of the important aspects of performance is the distribution of the accumulated gains across various agents. Fairness is an appropriate measure of evaluating the resource allocation schemes in such scenarios.

In many such problems, the resource allocated might be common across the agents but the utility derived by an agent depends upon the agent. Further there can be random variations in the utility derived based on the random state of the agent, at the time the resource is allocated. For example, consider a wireless communication system with a number of agents (mobiles) competing for resources from a single base station. The common resource to be allocated could be a time slot, while the utility derived could be the amount of information conveyed during the time slot, which in turn depends upon the state of channel between the agent allocated and the base station. Opportunistic resource allocation takes advantage of the time variations in the states of the competing agents and allocates resources to the users with 'best' state. This approach is shown to be advantageous when the random time variations in states of the agents are independent across the various time slots at which the allocation decisions are made. This approach results in an 'efficient' solution (see for e.g., [5], [11] and references therein), wherein the overall utility accumulated, because of such allocations, over all the time slots and over all the agents would be the maximum. This is a desirable feature from the point of view of the central unit (CU) allocating the resources. This however can regularly deprive the agents, whose states are 'bad' with high probability, resulting in very small or almost negligible utility accumulations for them. This is not acceptable from the point of view of the starved agents and that lead to the notion of fairness.

Fairness is a well studied concept (see [5], [4], [3], [11] and references therein). The notion of proportional fairness is introduced by Kelly et.al. [1] and a recent survey of various aspects related to fairness is available in [2]. Fairness is studied either via Jain's index ([3]) or is achieved by the CU optimizing a certain concave function of the accumulated utilities called α-fair function ([5], [11], [4]). Well known notions of fairness, namely proportional fairness, max-min fairness, etc., are achieved in this manner. A recent work ([4]) develops a family of fairness functions, parametrized by two numbers, which unify all the previously developed fairness measures. They show that this is the only family of functions satisfying four simple axioms of fairness metrics.

In this paper, we focus on a completely different view point related to the same problem. The central question asked here is "How much exactly is the improvement of the accumulated utilities of the starved users because of fair allocation ?" We obtain some answers in a very simple scenario, when one uses α-fair schedulers ([5], [11]).

Few answers

Let there be M agents with $m \in \{1, \cdots, M\}$ being a typical agent. Let $\mathbf{u} = (u_1, \cdots, u_M)$ represent the vector of (random) instantaneous utilities of all the M agents and let $\beta = (\beta_1, \cdots, \beta_M)$ represent the scheduler in the following sense: $\beta_m(\mathbf{u})$ represents the probability with which agent m is allocated given that the current vector of instantaneous utilities equals $\mathbf{u}$. An α-fair resource allocation β is obtained by maximizing the following function ([5], [11]):

$$\max_\beta \sum_m \Gamma^\alpha(\bar{u}_m(\beta)) \text{ with } \bar{u}_m(\beta) := E[u_m \beta_m(\mathbf{u})],$$

$$\Gamma^\alpha(\bar{u}) := \frac{\bar{u}^{1-\alpha}\mathbb{1}_{\{\alpha \neq 1\}}}{1-\alpha} + \log(\bar{u})\mathbb{1}_{\{\alpha=1\}}.$$

Here $\bar{u}_m(\beta)$ is the accumulated utility of agent m and by [11, Lemma 1] the above α-fair scheduler satisfies the following:

$$\bar{u}_{m;\alpha} = E\left[u_m \Pi_{j\neq m} \mathbb{1}\left\{\frac{u_m}{(\bar{u}_{m;\alpha})^\alpha} \geq \frac{u_j}{(\bar{u}_{j;\alpha})^\alpha}\right\}\right].$$

We first consider an extremely simplified scenario and obtain the following result about the accumulated utilities $\{\bar{u}_{m;\alpha}\}$ (proof in Appendix A).

Theorem 1. *Consider a scenario with instantaneous rates given by $u_m = a_m X_m$, where $\{X_m\}_{m \le M}$ are IID random variables and $a_m > 0$ is a constant for every m. That is, the instantaneous utilities of the agents are identically distributed except for constant multipliers.*

i) When a proportional fair scheduler (when $\alpha = 1$) is used the accumulated utilities are proportional to their 'mean value' i.e., to a_m:

$$\bar{u}_{m;1} = a_m \bar{u}_c \text{ with constant } \bar{u}_c := E[X_m \Pi_{j \neq m} \mathbb{1}_{\{X_m \ge X_j\}}].$$

ii) With two agents (i.e., $M = 2$), the one with larger a_m (w.l.g. let $a_2 > a_1$) will have higher accumulated utility when an efficient scheduler ($\alpha = 0$) is used and this accumulated utility decreases while that of the lower utility agent (i.e., agent 1) increases as the fairness factor increases (i.e., as α increases):

a) $\bar{u}_{2;0} > \bar{u}_{1;0}$; b) As $\alpha \uparrow \infty$, $\bar{u}_{2;\alpha} \downarrow$ and $\bar{u}_{1;\alpha} \uparrow$. $\diamond$

In a rather simplified scenario of Theorem 1, we could predict some properties of the accumulated utilities as a function of fairness factor, α. We could probably extend some of these results little more, may be when restricted to the setting of Theorem 1 or probably under slightly weaker assumptions. But it not clear whether these answers can be obtained under fairly general assumptions.

Alternatively, these answers can directly be obtained, if fairness is achieved by demanding a lower bound on the accumulated utilities of the starved users. So, we propose a constrained optimization to achieve a 'fair' scheduler (see Section III). In general, this is an infinite dimensional linear optimization problem with linear constraints. We reduce this to a finite dimensional optimization problem and following are the intuitive ideas behind the same. The instantaneous efficient decisions (without considering fairness) always chose the one with maximum instantaneous utility. It is clear that, to achieve fairness, one needs to deviate from the efficient decisions for some values of instantaneous utilities. It is also intuitive that we obtain the best, even under the 'fairness' constraints, if these deviations from the efficient decisions are made when the deviations result in 'minimal' losses. In [9], [10], Kleinrock et. al. introduced a class of priority based scheduling algorithms in the context of queuing systems with multiple classes of customers, wherein a scheduling decision is made after multiplying the waiting time of the longest waiting customer of each class with a priority factor assigned to the class. We observe that a similar *parametrized class of schedulers can achieve the required 'monotone' shift of efficient decisions and prove that achieving fairness is equivalent to choosing an appropriate priority factor.*

We make precise these concepts in this paper and show further that the fairness constraint optimization can be solved using a very simple iterative zero finding algorithm, whose complexity is similar to the complexity of the original α fair scheduling algorithms of [5].

In Section II the problem is introduced. In Section III one family of fair schedulers are introduced while the other extensions and comparisons are discussed in Section V. Conclusions and future directions are also discussed in the same section. Numerical examples are provided in Section IV and the proofs and related details are provided in Appendices.

II. System and background

A common resource is shared by M agents. Upon allocation of the resource, agent m gains a utility u_m, where u_m is completely determined by its state h_m. The state of the resource h_m is a random variable and is independent across the agents. The time is divided into slots and resource is allocated once in a slot. We are interested in the accumulated utilities, or equivalently the time averages of the utilities and we study the same via their expected values. The state h_m for any agent m is an identical and independent process across the time slots. We assume that the instantaneous utilities $\{u_m\}$ are deterministic functions of the states $\{h_m\}$ and to simplify the notations we work directly with the random utilities $\{u_m\}$. We assume $u_m \in \mathcal{U}_m$, where $\mathcal{U}_m$ is a subset of $\mathcal{R}$. The agents are interested in the average utility obtained by themselves:

$$\bar{u}_m(\beta) = E[u_m \beta_m(\mathbf{u})] \text{ with } \mathbf{u} := (u_1, u_2, \cdots, u_M).$$

The aim of a central unit (CU) is to obtain a scheduler $\beta = (\beta_1, \cdots, \beta_M)$ which achieves a 'given goal' with respect to these average utilities.

This resource allocation problem is a typical example of a multi agent game theoretic problem in which the central unit (CU) also participates. The utility of an agent competing for resources is obvious: it is the utility accumulated by the agent due to allocations ($\bar{u}_m$ for agent m). On the other hand the utility of central unit (CU) in the game theoretic context will depend upon its objective, like for example, fairness-efficiency trade-off, level of fairness, etc., (see [11] for more details). We do not pursue the game theoretic approach in this paper, rather transfer the goals of the individual agents to the CU itself. It is the CU which ensures that the 'fairness' goals of all the agents are satisfied, whenever possible as would be evident in the sections below.

Efficient scheduler

As already mentioned the CU can have varying objectives depending upon the design. If its gain is based solely upon the total utility transferred, irrespective of the agent that received the utility, it is would naturally schedule in every time slot the agent with maximum instantaneous utility. This is a situation in which the CU maximizes the sum of the average utilities over an appropriate domain:

$$\max_{\beta \in \Omega} \Upsilon(\beta)$$
$$\text{with } \Upsilon(\beta) := \sum_m \bar{u}_m(\beta) = E\left[\sum_m u_m \beta_m(\mathbf{u})\right]. \qquad (1)$$

We next describe the domain of optimization Ω. Let $\|.\|_2$ represent the L^2-norm, i.e., $\|u_m\|_2 = \sqrt{E[u_m^2]}$. We extend this naturally to any $\beta = (\beta_1, \cdots, \beta_M) \in L^2$ (we still denote this space by L^2) by

$$\|\beta\|_2 = \sqrt{\sum_m \|\beta_m\|_2^2} = \sqrt{\sum_m E[\beta_m^2]} \text{ for any } \beta \in L^2.$$

Now the domain of optimization $\Omega \subset L^2$, is defined as below:

$$\Omega := \{\beta : \Pi_m \mathcal{U}_m \to [0,1]^M \, ; \, \sum_m \beta_m(\mathbf{u}) = 1 \text{ for all most all } \mathbf{u}\}. \quad (2)$$

The scheduler maximizing (1) is called an efficient scheduler and it is easy to see in this case that the efficient scheduler is given by

$$\beta_m^{eff}(\mathbf{u}) = \mathbb{1}_{\{\arg\max_k u_k = m\}}. \quad (3)$$

Let $\bar{u}_m^{eff}$ denote the corresponding accumulated utilities, i.e.,

$$\bar{u}_m^{eff} := \bar{u}_m(\beta^{eff}) = E[u_m \mathbb{1}_{\{\arg\max_k u_k = m\}}]. \quad (4)$$

III. Fairness via constrained optimization

Consider a simple example scenario with two agents. The states of one agent is better than the other with probability one, i.e., assume $u_1 > u_2$ with probability one. In this case, if CU uses efficient scheduler (3), then agent 2 is never allocated a resource and its average utility is zero. Similarly, if an agent has its utility lesser than the other agent with high probability, he would obtain non zero utility, but nevertheless would be a very small quantity. So, such agents are discouraged to use the facility offered by the CU. This calls for a different scheduler which ensures that all the agents obtain 'satisfactory' utility. That is, the scheduler should ensure every agent obtains a fair share for himself. The "fair shares" could depend upon

1) the context (e.g., in wireless communications, user with data transfer requests can wait longer than the user with a communication call request),
2) the price paid for service (e.g., one class of users might be willing to pay more to obtain better service),
3) type/size of requests (e.g., some users might require short jobs while others might require service for longer duration) etc.

When one deviates from the efficient scheduler (1) to obtain a required level of fairness, it is clear that efficiency is reduced. So, the goal of this work is to introduce a class of schedulers which maximize the efficiency under the constraint that fairness is maintained to a 'required level'.

We begin by introducing the following notion of Θ-fairness (with a given $\Theta := (0, \theta_2, \theta_3, \cdots, \theta_M)$), given by the following constrained optimization

$$\beta_\Theta^* \;\; := \;\; \arg\max_{\beta \in \Omega} \Upsilon(\beta) \;\; \text{with} \;\; \Upsilon(\beta) := \sum_m \bar{u}_m(\beta)$$
$$\text{subject to } \bar{u}_m(\beta) \geq \theta_m \text{ for all } m \geq 2. \quad (5)$$

Here without loss of generality, we assume $\bar{u}_1^{eff} \geq \bar{u}_m^{eff}$ for all $m \geq 2$.

In general, problem (5) is an infinite dimensional constrained optimization problem and it is difficult to solve the same. However, we show that a finite dimensional scheduler optimizes the above problem and towards this, we consider the following sub family of schedulers inspired by delay priority schedulers ([9]):

$$\Omega' := \{\beta(\mathbf{b}) = (\beta_1(\mathbf{b}), \beta_2(\mathbf{b}), \cdots, \beta_M(\mathbf{b})) \text{ with } \beta_m(\mathbf{b})(\mathbf{u}) := \mathbb{1}_{\{\arg\max_k u_k b_k = m\}}\}. \quad (6)$$

The intuitive reasons for this choice is already discussed in Introduction. As seen from (6), the family of priority schedulers are parametrized by the finite dimensional vector $\mathbf{b} = (b_1, b_2, \cdots, b_M)$. Throughout we assume $b_1 = 1$ while $b_m \geq 1$ for all m, which obviously improves the utilities of lower utility agents.

Assumptions

Without loss of generality (w.l.g.), we assume that the agents are arranged in the increasing order of their efficient utilities, i.e., $\bar{u}_1^{eff} \geq \bar{u}_2^{eff} \cdots \geq \bar{u}_M^{eff}$. We assume a work conserving principle, i.e., the CU always schedules the resource whenever there is a utility to be transferred (i.e., $\sum_m \beta_m = 1$ for all states with non zero utilities, i.e., whenever $\mathbf{u} \neq 0$) . We work under the following assumptions.

A.1 Assume $Prob(u_m \leq 0) = 0$ and $E[u_m^2] < \infty$ for all $1 \leq m \leq M$.
A.2 When $\theta_m > 0$, we assume $\theta_m > \bar{u}_m^{eff}$ for all $m > 1$.
A.3 The set of constraints in problem (5) are achievable, in fact there exists at least one $\beta \in \Omega$ such that: $\bar{u}_m(\beta) > \theta_m$ for all m with $\theta_m > 0$.
A.4 The instantaneous utilities u_m are continuous random variables for each m, i.e., they have density. These processes are also independent across the agents.

We immediately have the following two results, whose proofs are in Appendix A.

Theorem 2. *There exists a solution for problem (5) in Ω, under the assumption* **A**.*1.* $\diamond$

Theorem 3. *The optimizer of the constrained optimization (5) satisfies all the constraints under* **A**.*1 and* **A**.*2.* $\diamond$

We now consider two different cases: one when the utilities are continuous random variables (which satisfies **A**.4) and one when they are discrete. We will see that Ω itself is finite dimensional in the discrete case, while this is not true in the continuous case, but however a reduction is possible.

A. Continuous utilities

We begin with reducing the infinite dimensional constrained optimization problem (with domain Ω) to that of a finite dimensional one (Ω'). We also establish the uniqueness of the optimizer in Ω'.

Theorem 4. *Assume* **A**.*1-4. Then,*

1) There exists a $\beta^* \in \Omega'$ which is an optimizer of the constrained optimization problem (5), i.e., there exists a $\mathbf{b}^* \in R^{M-1}$ and:

$$\beta^* = \beta(\mathbf{b}^*), \text{ with } \beta_m(\mathbf{b}^*)(\mathbf{u}) := \mathbb{1}_{\{\arg\max_k u_k b_k^* = m\}}.$$

By Theorem 3, $\mathbf{b}^*$ *satisfies the fairness constraints with equality, i.e.,*

$$\bar{u}_m(\beta^*) = \theta_m \text{ for all } m \geq 2.$$

2) *Further for any* $m \geq 2$, $b^*_m = 1$ *if* $\theta_m = 0$, *i.e., for those agents without constraint.*
3) *There exists a unique element in* Ω' *which satisfies all the constraints. This unique element is the unique optimizer in* Ω' *for the constrained optimization problem (5).*

Proof: is in Appendix A. ◇

A new fair scheduler : We propose the following two time scale stochastic approximation based algorithm to obtain the zeros of fairness constraints equation. Here k represents the time slot.

$$\begin{aligned}
\bar{u}_{m,k+1} &= \bar{u}_{m,k} + \mu_k \left(u_{m,k} \mathbb{1}_{\{\beta_k = m\}} - \bar{u}_{m,k} \right) \\
&\qquad \text{for all } m \geq 1, \\
\beta_k &= \arg\max_m (b_{m,k} u_{m,k}), \\
b_{m,k+1} &= b_{m,k} + \epsilon_k (\bar{u}_{m,k} - \theta_m) \\
&\qquad \text{for all } m \text{ with } \theta_m > 0,
\end{aligned} \tag{7}$$

and with $b_{m,k} = 1$ for all m with $\theta_m = 0$. The first iteration estimates the average utilities of agent m, while the second iterate updates the scheduler representative $\mathbf{b}$ to converge towards a point that satisfies the constraints and hence is an optimal scheduler. The second iterate is updated at a slower rate:

$$\epsilon_k = C \mu_k \mathbb{1}_{\{k \in \{N, 2N, 3N, 4N, \cdots\}\}}$$

for some large N and suitable $C \geq 1$. In Section IV we establish that the above algorithm indeed obtains the required zeros asymptotically via some numerical examples.

One can alternatively consider a single time scale version, wherein the second iterate is updated fast and is updated directly with the help of the 1-estimate $u_{m,k} \mathbb{1}_{m=\beta_k}$, in place of $\bar{u}_{m,k}$ (for all m with $\theta_m > 0$):

$$b_{m,k+1} = b_{m,k} + \mu_k (u_{m,k} \mathbb{1}_{\{\beta_k = m\}} - \theta_m). \tag{8}$$

This algorithm is also considered in Section IV. The theoretical evidence to establish the convergence of these algorithms and choosing the best among the possible variants, would be taken up as a future work.

B. Discrete Utilities

Consider a case in which all agents have discrete utilities. Agent m can take any one of N_m different utilities and let $\bar{N} := \Pi_{m \leq M} N_m$ define the total possible number of instantaneous utility vectors $\mathbf{u}$. By indexing each one of $\mathbf{u}$ equivalently with a number $1 \leq j \leq \bar{N}$, one can rewrite the domain of optimization as a subset of $[0,1]^{\bar{N}M}$ as below:

$$\Omega = \Big\{ \beta = \{(\beta_1(j), \cdots, \beta_M(j))\}_{j \leq \bar{N}} \text{ with } \beta_m(j) \in [0,1], \\ \forall m, j \text{ and such that } \textstyle\sum_m \beta_m(j) = 1 \text{ for all } j \leq \bar{N} \Big\}.$$

Interesting point to notice for discrete case is that, Ω by itself is a finite dimensional subset. However in this case, many conclusions of Theorem 4 are not true and the technical reasons for the same is discussed in Appendix A after the proof of Theorem 4. We mainly do not have some uniqueness properties and in Appendix B, via a simple example we discuss the issues and advantages related to discrete case. However Theorems 2, 3 are true and by virtue of these one can again obtain the optimizer as a (finite dimensional) zero of "fairness constraints equation". One can modify the algorithm (7), in an obvious way, to update directly the scheduler components $\{\beta_m(j)\}$.

$$\begin{aligned}
\bar{u}_{m,k+1} &= \bar{u}_{m,k} + \mu_k \left(u_{m,k} \mathbb{1}_{\{\beta^a_k = m\}} - \bar{u}_{m,k} \right) \\
&\qquad \text{for all } m \geq 1, \\
\beta^a_k &= \text{ a random allocation generated with} \\
&\qquad \text{probability vector } (\beta_{1,k}(\mathbf{u}_k), \cdots, \beta_{M,k}(\mathbf{u}_k)) \\
\beta_{m,k+1}(\mathbf{u}_k) &= \beta_{m,k}(\mathbf{u}_k) + \epsilon_k (\bar{u}_{m,k} - \theta_m) \\
&\qquad \text{for all } m \text{ with } \theta_m > 0 \text{ and} \\
\beta_{m^*_k,k+1}(\mathbf{u}_k) &= 1 - \sum_{m:\theta_m > 0} \beta_{m,k+1}(\mathbf{u}_k); \text{ with} \\
m^*_k &= \arg\max_m u_{m,k}, \text{ the efficient decision.}
\end{aligned} \tag{9}$$

Of course the required projection has to be take care in updates of $\{\beta_{m,k}\}$. However we do have an issue: not all the zeros are optimizers (see Appendix B). Further the dimension of Ω increases with the number of states and or the number of agents and this increases the computational complexity. In such cases, one can approximate the optimal scheduling policy with that of an appropriate 'continuous counterpart' and this procedure is explained in the technical report ([12]). Since there exists a unique zero in Ω' in continuous case which is also the maximizer, we expect this to represent approximately the optimizer of the discrete case also, whenever the dimension is large.

IV. Illustrative Examples

A. Continuous utilities: Example 1

By Theorem 4, Θ-fairness is achieved by an unique zero of the equation representing fairness constraints in the finite dimensional Ω'. We begin with a very simple example of uniformly distributed utilities and obtain the required unique zero, and thereby achieve the required Θ-fair scheduling, using the single time scale algorithm (8).

We consider a three agent example, with agents having utilities $\{u_m\}$ that are uniformly distributed between 0 and $u_{max} = [4, 1.8, 1]$. The efficient utilities of the three agents, for this example, can easily be estimated and they equal [1.946, 0.245, 0.027]. We can see that the agents 2 and 3 are starved and we would like to obtain an $\Theta = [0, 0.6, 0.1]$ scheduler in Figure 1. We notice that the average utilities $\bar{u}_{m,k}$ as well as the scheduler representatives $b_{2,k}$, $b_{3,k}$ converge within 4000 iterations to satisfy the fairness constraints. For larger systems (as considered in the coming examples) the convergence might be slower, but would still

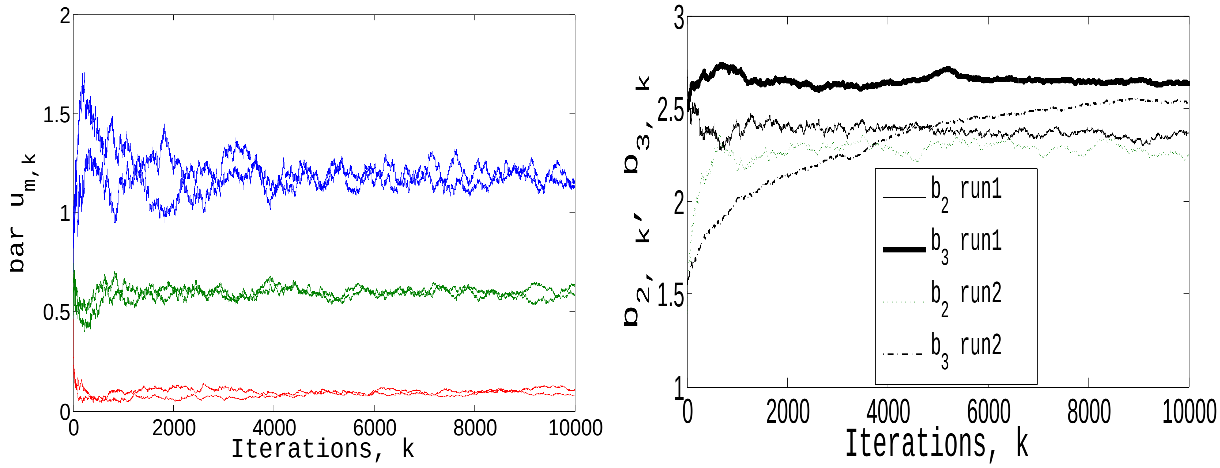

Fig. 1. Example 1 (Uniform utilities): Two simulation runs with different initial conditions. Convergence of iterates $\bar{u}_{m,k}$ in the left figure (satisfying the constraints) and Convergence of iterates $[b_{2,k}, b_{3,k}]$ in right figure. With single time scale algorithm, $\bar{u}_{m,k}$ is used only for illustrative purposes.

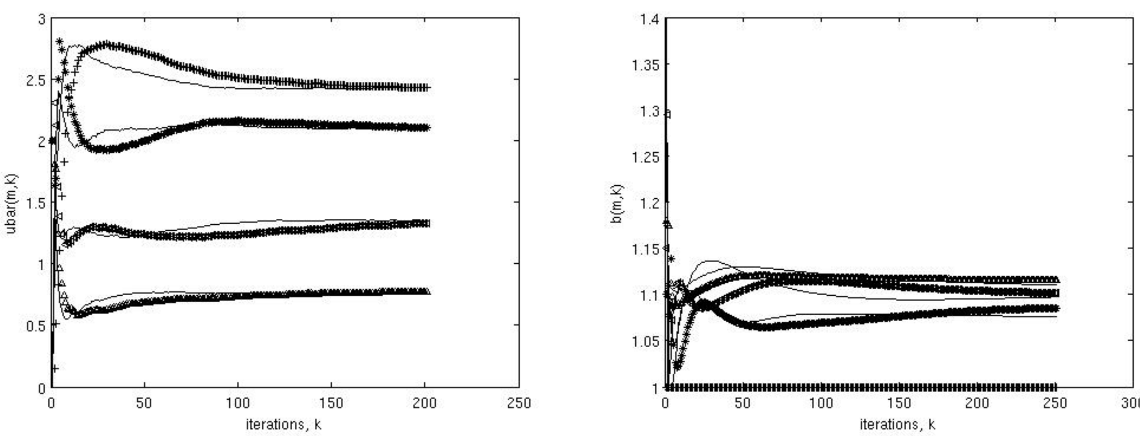

Fig. 2. Example 2 (Rayleigh states): Two simulation runs with different initial conditions. Convergence of iterates $\bar{u}_{m,k}$ in the left figure and Convergence of iterates $\mathbf{b}_k$ in right figure

Initial b	**b converges to**	**Final utilities**
1 2.00 2.00 2.00	1. 1.087 1.127 1.135	2.496 2.103 1.299 0.736
1 1.50 1.50 1.50	1. 1.083 1.112 1.126	2.424 2.085 1.350 0.772
1 2.00 1.50 2.00	1. 1.084 1.122 1.137	2.506 2.120 1.251 0.754
1 2.10 1.90 1.60	1. 1.085 1.130 1.133	2.462 2.106 1.300 0.760
1 1.20 1.30 1.40	1. 1.076 1.097 1.110	2.491 2.106 1.277 0.768
1 1.10 1.15 1.18	1. 1.085 1.101 1.115	2.492 2.081 1.315 0.753
1 1.40 1.40 1.40	1. 1.081 1.107 1.120	2.434 2.093 1.346 0.761
1 1.80 1.80 1.80	1. 1.081 1.125 1.133	2.431 2.099 1.326 0.772

TABLE I
CONTINUOUS UTILITIES EXAMPLE 2: DIFFERENT INITIAL CONDITIONS, BUT CONVERGENCE TO THE SAME LIMIT

be comparable to that of the existing fair schedulers ([5]). With two time scale version (7), the convergence is much slower.

B. Continuous utilities: Example 2

We consider the context of wireless communications and look at the following problem: M cellphones are competing for resource from a single central unit or base station. In this context, h_m represents the channel state between the base station and user m. For any agent m, h_m is a continuous random variable and $\{h_m\}$ are independent across the agents and independent and identical across the time slots. We assume that h_m is a Rayleigh random variable. The maximum rate (capacity) at which data can be transferred over the communication channel is the utility u_m, derived by agent m when allocated the resources. This (capacity) utility u_m is a deterministic function of h_m and is given by,

$$u_m = log(1 + h_m^2).$$

We consider the case with 4 agents ($M = 4$) and the Rayleigh parameters of the 4 agents are given by [20, 15, 12, 10]. The goal here is to see

1) The convergence of expected utilities for different agents.
2) Existance of unique zero of the "fairness constraints equation", i.e., convergence of vector $\mathbf{b}$ irrespective of the initial condition.

The efficient scheduler results in the following accumulated allocations:

$$\bar{\mathbf{u}}^{eff} = [3.618, 1.784, 0.867, 0.426].$$

The fairness constraint vector Θ is set as,

$$\Theta = [0, 2.1, 1.3, .75].$$

The other parameters of the two-time scale algorithm (7) are

$$\mu_k = 1/k + 0.01/k^3,\ N = 100 \text{ and } C = 50.$$

That is, fairness parameters $\{b_{m,k}\}$ are updated once in 100 steps. For the example at hand, we found these to be a good set of parameters. Rayleigh distributed random variable h with parameter σ is generated from uniformly distributed random variable U via the following transformation:

$$h = \sigma\sqrt{-2logU}.$$

Observations

Figure 2 illustrates the convergence behavior of the iterates of algorithm (7) while Table I studies the limits of the same algorithm. In the left figure one sample in 250 samples is plotted while in the right figure one in 200 samples is plotted. This is done for the clarity of the figures. From table, it can be seen that the expected utilities converge and all the fairness constraints are satisfied within a negligible margin of error, irrespective of the initial conditions, as established by Theorem 4. Thus this algorithm can be used to obtain the required zero and thereby to implement the Θ-fairness for any given Θ.

C. Discrete case

We now consider an example with discrete utilities i.e. the case when **A**.4 does not hold. Here Ω itself is a finite dimensional set but the computational complexity increases as the size $\bar{N}$ (see Section III-B) increases. We present here a method, wherein the Θ-fair scheduler of the discrete scenario can be well-approximated by the scheduler of an appropriate continuous case and this approximation is good when the cardinality $\bar{N}$ is large. We continue with the example of wireless communication problem from the continuous case study, but with a difference. Here the system supports only a finite number of transfer rates, say cellphone m can be supported by one of the rates given

$$\mathcal{U}_m = \{0, log(1+\delta_m^2), log(1+(2\delta_m)^2), \cdots, log(1+(N_m\delta_m)^2)\},$$

where $\delta_m > 0$. The channel state h_m is Rayleigh as in previous example, but now the utility of the mobile m is

$$u_m = \max\{u \in \mathcal{U}_m : u \leq log(1 + h_m^2).\}$$

Clearly as $(\delta_m, N_m) \to (0, \infty)$, $u_m \to log(1 + h_m^2)$ for almost all h_m and we approximate such cases with the continuous case of the previous example.We consider one such case in which all the agents are having equal number of transfer rate choices, i.e, $N_m = N_1$ for all m. In this experiment we calculated $\{\delta_m\}$ such that the Rayleigh density of every mobile at $\delta_m N_1$ (every channel state after this value is supported by the common highest rate) equals 90 percentile value. All the other parameters are same as in the previous example, including Θ.

We again use the algorithm (7), even when it is a discrete case. The results are tabulated in Table II. Interestingly even up-to the case with $N_1 = 20$, all the constraints are (well) satisfied approximately, by an element from $\beta(\mathbf{b}) \in \Omega'$, to which the algorithm converges asymptotically. From Theorem 3 (which is true even for the case with discrete utilities) the optimizer satisfies the constraints. More interestingly, the scheduler representative $\mathbf{b}$ is close to the corresponding one in the continuous case. And we know in the continuous case that the (unique) zero of the constraints equation is the maximizer. Thus we expect the (approximate) zeros so obtained to be the Θ-fair schedulers even for the discrete case. We do notice from the same table that for the case with $N_1 = 5$, the constraints are not satisfied. Thus this method is not accurate, when the cardinality of the utility space is small. However in such cases the dimension of Ω itself is small and once can directly use the algorithm (9) indicated for the discrete case.

V. Extensions, comparisons and discussions

One can generalize and define f-fairness for M-agents as in the following. Let $f = (f_1, \cdots, f_M) : \mathcal{R}_+^M \to \mathcal{R}_+^M$ be any given measurable function and define, f-fair scheduler as:

$$\beta^f := \arg\max_{\beta\in\Omega} \sum_i \bar{u}_i(\beta)$$
$$\text{subject to } f_m(\bar{u}_1(\beta), \cdots, \bar{u}_M(\beta)) \geq 0 \text{ for all } m. \quad (10)$$

N_1	Initial **b**	**b** converges to	Final utilities
continuous	1, 1.4, 1.4, 1.4	1. 1.081 1.107 1.120	2.434 2.093 1.346 0.761
40	1, 1.4, 1.4, 1.4	1. 1.124 1.112 1.092	2.020 2.130 1.319 0.759
20	1, 1.4, 1.4, 1.4	1. 1.179 1.130 1.087	1.899 2.008 1.299 0.750
5	1, 1.4, 1.4, 1.4	1. 2.600 1.000 1.000	1.620 2.119 1.014 0.008

TABLE II
DISCRETE UTILITIES EXAMPLE: THE ALGORITHM (7) PERFORMS WELL FOR N_1 = 20 OR MORE

Note that $f_m(\mathbf{u}) = u_m - \theta_m$ gives the notion of Θ-fairness introduced earlier. One can explore other options via this general definition. For example, we are interested in the following function:

$$f_m(u_{m-1}, u_m) \quad = \quad \bar{u}_m - \gamma_m \bar{u}_{m-1} \text{ for all } m > 1.$$

Essentially we are interested in maintaining the ratio:

$$\frac{\bar{u}_m}{\bar{u}_{m-1}} \geq \gamma_m \text{ for all } m > 1.$$

The functions of this type can equivalently be represented using the vector $\Gamma := (1, \gamma_2, \cdots, \gamma_M)$ and the corresponding fairness can be termed as Γ- fairness.

All the proofs go through for this notion of fairness also. We just mention few differences. Theorem 3 will go through in a similar way if we assume as in Assumption **A**.2 that we aim to promise better returns for starved agents, which are better than what they could obtain using efficient scheduler, that is:

$$\gamma_m > \frac{\bar{u}_m^{eff}}{\bar{u}_{m-1}^{eff}} \text{ for all } m > 1.$$

In step 17 of Theorem 4 we will have

$$b_m^* = 1 + \varrho_m^* - \gamma_m \varrho_{m-1}^* \text{ for all } m > 1.$$

To prove Theorem 5 version for Γ-fairness, if possible consider two vectors $\mathbf{b}$, $\mathbf{c}$ both of which satisfy the ratio constraints and such that $\mathbf{b} \neq \mathbf{c}$. Then if w.l.g. $\bar{u}_1(\mathbf{b}) > \bar{u}_1(\mathbf{c})$, and then to maintain the ratios we will need recursively that $\bar{u}_m(\mathbf{b}) > \bar{u}_m(\mathbf{c})$ for all m. Which is not possible because of work conserving principle. When $\bar{u}_1(\mathbf{b}) = \bar{u}_1(\mathbf{c})$, then again recursively to maintain the ratios, we will have $\bar{u}_m(\mathbf{b}) = \bar{u}_m(\mathbf{c})$ for all m. In all, this means we might have $\mathbf{b} \neq \mathbf{c}$ but with $\bar{u}_m(\mathbf{b}) = \bar{u}_m(\mathbf{c})$ for all m to satisfy the ratio constraints. But this is immediately contradicted by the existing Theorem 5.

Comparison with existing notions of fairness

The unconstrained optimization (with $\Theta = (0, 0, \cdots, 0)$) gives efficient scheduler.

When $\Gamma = (1, \cdots, 1)$ we obtain the max min fairness as this ensures all agents have equal utility.

There exists a Γ which defines the proportional fair scheduler. For example, for the case considered in Theorem 1 given in the Introduction and using the same theorem, Γ defined by

$$\Gamma = \left(1, \frac{a_2}{a_1}, \frac{a_3}{a_2}, \cdots, \frac{a_{M-1}}{a_M}\right),$$

achieves proportional fairness. This is possible because for any given Γ there exists an unique vector of accumulated utilities. This is true because of the work conserving principle as explained in the previous paragraph while indicating the proof of Theorem 5 for the case of Γ-fairness. We can also achieve this via Θ-fairness as given below.

As mentioned in the introduction, every α-fair scheduler satisfies the following equation ([11, Lemma 1])

$$\begin{aligned} \beta_m^{*,\alpha}(\mathbf{u}) \quad &= \quad \mathbb{1}_{\left\{\arg\max_{m'} \left(\bar{u}_{m';\alpha}\right)^{-\alpha} u_{m'} = m\right\}} \text{ with} \\ \bar{u}_{m;\alpha} \quad &= \quad E[u_m \beta_m^{*,\alpha}(\mathbf{u})]. \end{aligned}$$

This implies α-fair scheduler is obtained by the priority schedulers of Section III with priority factors

$$b_m = \left(\frac{\bar{u}_{1;\alpha}}{\bar{u}_{m;\alpha}}\right)^{\alpha} \text{ for all } m \geq 2.$$

For every α there exists a Θ^α such that the Θ^α scheduler of this paper obtains the corresponding α fair scheduler and in fact $\Theta^\alpha = [0, \bar{u}_{2;\alpha} \cdots, \bar{u}_{M;\alpha}]$.

Conclusions and Future directions

In this paper, we introduced a notion of fairness which provides direct information about the improvement of the accumulated utilities of the otherwise starved agents. Via a series of theorems we proved that this type of fairness can be achieved by algorithms whose complexity is similar to those already proposed in literature for the other notions. The notion introduced here has some connections to the research done before. For example, in wireless communications which has to support both data calls (lengthy connections which can wait) and voice calls (impatient and short calls, which drop-off if all the servers are busy), it is proposed to optimize the average waiting time of a data call while ensuring that the drop probability of a voice call is below certain acceptable limit.

We also introduced Γ-fairness which ensures that the ratios of utilities of any pair of agents is better than that achievable using efficient scheduler. Using this we can achieve the existing notion of max-min fairness. In an extremely simplified scenario (Theorem 1), we showed that with a Proportional fair scheduler the accumulated utilities are proportional to their means. Once the mean of the instantaneous utilities ($E[u_1]$, $\cdots$, $E[u_M]$) of all the agents are known (one can easily estimate the mean), one can actually implement a fair scheduler which ensures that the accumulated utilities of various agents are proportional to their own mean values.

For example, one can achieve this using Γ-fairness defined in this paper for

$$\Gamma := \left(1, \frac{E[u_2]}{E[u_1]}, \cdots, \frac{E[u_{M-1}]}{E[u_M]}\right).$$

We would like to complete this analysis and investigate more on these aspects in future.

In this paper we indicated few iterative algorithms to implement the new notion of fairness. We would like to study their convergence properties. In view of Theorem 4 alternatively, one may obtain this as the optimizer in Ω' and design an iterative algorithm to achieve the same. The case with discrete utilities needs to be explored more.

REFERENCES

[1] F. Kelly, A. Maulloo, D. Tan, "Rate control for communication networks: shadow prices, proportional fairness, and stability", Journal of the Operations Research Society 49 (1998).

[2] Wierman, Adam. "Fairness and scheduling in single server queues." Surveys in Operations Research and Management Science 16.1 (2011): 39-48.

[3] R. Jain, D. M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system", Eastern Research Laboratory, Digital Equipment Corp., 1984.

[4] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, An axiomatic theory of fairness in network resource allocation, in Proceedings of IEEE INFOCOM. IEEE, 2010, pp. 19.

[5] H.J. Kushner, P.A. Whiting, "Convergence of Proportional-Fair Sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.* vol. 3, no. 4, pp. 12501259, Jul. 2004.

[6] V. K. N. Lau, "Proportional Fair Space Time Scheduling for Wireless Communications", IEEE Trans. communications, vol. 53, no. 8, Aug 2005.

[7] D.G. Luenberger, "Convex Programming and Duality in Normed Space" IEEE Trans. on System Science and Cybernetics, July 1968.

[8] D.G. Luenberger, "Optimization by vector space methods", John Wiley and Sons, Inc., 1969.

[9] Leonard Kleinrock, and Roy P. Finkelstein. "Time dependent priority queues." Operations Research 15.1 (1967): 104-116.

[10] Leonard Kleinrock, "A delay dependent queue discipline", Naval Research Logistics Quarterly, volume 11, pages 329-341, 1964.

[11] Veeraruna Kavitha, Eitan Altman, Rachid Elazouzi, Rajesh Sundareshan, "Fair scheduling in cellular systems in the presence of noncooperative mobiles" accepted at IEEE Trans. on Networking.

[12] Veeraruna Kavitha, N. Hemachandra, Debayan Das, "Fairness via priority scheduling", technical report downloadable using http://www.ieor.iitb.ac.in/files/faculty/kavitha/FairPriority.pdf

APPENDIX A

This appendix contains all the proofs of the theorem given in the main body of the paper. It also contains the theorems and lemmas used for establishing those theorems.

Proof of Theorem 1: The first part is easily evident by substituting $\{a_m \bar{u}_c\}_m$ into the fixed point equation given above the theorem. For the next two parts, we have the case with $M = 2$. When $\alpha = 0$, because $X_1 \stackrel{d}{=} X_2$:

$$\begin{aligned}
\bar{u}_{2;0} &= a_2 E[X_2 \mathbb{1}_{\{a_2 X_2 > a_1 X_1\}}] \\
&= a_2 E[X_1 \mathbb{1}_{\{a_2 X_1 > a_1 X_2\}}] \\
&> a_2 E[X_1 \mathbb{1}_{\{a_1 X_1 > a_2 X_2\}}], \text{ because } a_2 > a_1, \qquad (11) \\
&> a_1 E[X_1 \mathbb{1}_{\{a_1 X_1 > a_2 X_2\}}], \text{ because } a_2 > a_1, \\
&= \bar{u}_{1;0}.
\end{aligned}$$

Define $\xi^\alpha := a_2 (\bar{u}_{1;\alpha})^\alpha / a_1 (\bar{u}_{2;\alpha})^\alpha$ and define

$$\begin{aligned}
\nu_1^\alpha &= E[X_1 \mathbb{1}_{\{X_1 > \xi^\alpha X_2\}}] \text{ and} \\
\nu_2^\alpha &= E[X_1 \mathbb{1}_{\{X_1 \leq \xi^\alpha X_2\}}]. \qquad (12)
\end{aligned}$$

Clearly either $\nu_1^\alpha \uparrow$ and $\nu_2^\alpha \downarrow$ with α or the vice-versa (note $\nu_1^\alpha + \nu_2^\alpha$ is a constant for all α). Note (as in previous step) by fixed point equation that, $\bar{u}_m^\alpha = a_m \nu_m^\alpha$ for all m and hence that $\xi^\alpha = (a_1/a_2)^{\alpha-1}(\nu_1^\alpha/\nu_2^\alpha)^\alpha$. When $\alpha = 0$, while proving the first part of theorem we see that $\nu_2^0 \geq \nu_1^0$ (see inequality 2 in equation (11)).

Now if $\nu_2^\alpha \uparrow$ and $\nu_1^\alpha \downarrow$ with $\alpha \uparrow$, then $\nu_2^\alpha/\nu_1^\alpha \uparrow$ and then $\nu_2^\alpha/\nu_1^\alpha > 1$ for all $\alpha \geq 0$ and hence we also have $\xi^\alpha \downarrow$. But then from (12) $\nu_1^\alpha \uparrow$ which is a contradiction. Thus, $\nu_1^\alpha \uparrow$ and $\nu_2^\alpha \downarrow$ with $\alpha \uparrow$. ◇

Proof of Theorem 2: Clearly Ω is a convex subset of L^2 and L^2 is a Hilbert space. Consider a smaller subset of Ω which defines the constraints of (5) as below:

$$\Omega_\theta = \{\beta \in \Omega : \bar{u}_m \geq \theta_m \text{ for all } m \geq 2\}.$$

By linearity of the map $\beta \mapsto \bar{u}_m(\beta)$, Ω_θ is also a convex subset.

Further it is a closed subset. To see that, if there exists a subsequence $\{\beta^n = (\beta_1^n, \cdots, \beta_M^n)\} \subset \Omega_\theta$ with $\beta^n \to \beta$ in L^2 norm. Then, $\beta^n \to \beta$ in probability and this implies the existence of a subsequence along which $\beta^{n_k} \to \beta$ almost surely. The almost sure convergence gives us that:

$$1 \leq \beta(\mathbf{u}) \leq 1 \text{ and } \sum_m \beta_m(\mathbf{u}) = 1 \text{ for almost all } \mathbf{u}.$$

Further by dominated convergence theorem $\bar{u}_m(\beta^{n_k}) \to \bar{u}_m(\beta)$. Thus $\bar{u}_m(\beta) \geq \theta_m$ for all m, and so $\beta \in \Omega_\theta$ and hence Ω_θ is closed.

The function $\beta \mapsto \Upsilon(\beta) = \sum_m \bar{u}_m(\beta)$ is clearly linear and is bounded because by Cauchy-Schwartz inequality:

$$\begin{aligned}
\|\Upsilon\| &= \sup_{\|\beta\|_2 \leq 1} |\sum_m \bar{u}_m(\beta)| \leq \sup_{\|\beta\|_2 \leq 1} \sum_m \|\beta_m\|_2 \|u_m\|_2 \\
&\leq \sum_m \|u_m\|_2 < \infty.
\end{aligned}$$

Thus, Υ is a continuous linear functional and and hence is Gateaux differentiable function. Thus[1], there exists a $\beta^* \in \Omega_\theta$ which maximizes Υ.

◇

Proof of Theorem 3: If say there exists a $\beta^* = (\beta_1^*, \cdots, \beta_M^*)$ which maximizes the optimization problem (5) without satisfying the constraint with equality. That is to say for some m with $\theta_m > 0$,

$$\int u_m \beta_m^*(\mathbf{u}) dP(\mathbf{u}) - \theta_m = \delta > 0.$$

[1] We have the following theorem from Functional analysis.

Theorem. *Let K be a nonempty closed convex set of a Hilbert space V, and let $J : V \to R$ be a convex Gateaux differentiable function. If K is bounded, or if J is infinite at infinity, there exists at least one minimum of J over K, i.e.,*

$$J(u) = inf_{v \in K} J(v) \text{ for some } u \in K.$$

By **A**.2, $\theta_m > \bar{u}_m^{eff}$ given by (4) and this implies, agent m obtains more because of 'non efficient' scheduling decisions (ones with $\beta_m^*(\mathbf{u}) > 0$ when $m \neq \arg\max_k u_k$), i.e., $P(A) > 0$ where

$$A := \left\{ \mathbf{u} : \beta_m^*(\mathbf{u}) > 0 \text{ when } u_1 = \max_k u_k \right\}.$$

The reason for the above is: the 'non efficient' decisions are obtained by shifting the 'efficient' decisions of only the un constrained agents (agents with $\theta_m = 0$) in favor of the agents with constraints on non zero measure set and there should be at least one such agent, assume without loss of generality, this to be the agent 1.

This implies there exists a $\tilde{\delta}_1 > 0$, $\tilde{\delta}_2 > 0$ and a set $B_{\tilde{\delta}_1,\tilde{\delta}_2} \subset A$ such that[2] $P(B_{\tilde{\delta}_1,\tilde{\delta}_2}) > 0$ where

$$B_{\tilde{\delta}_1,\tilde{\delta}_2} := \left\{ \beta_m^*(\mathbf{u}) > \tilde{\delta}_1 \text{ with } \max_k u_k = u_1 \text{ and } u_1 - u_m > \tilde{\delta}_2 \right\}.$$

Then a new scheduler policy is defined as the following:

$$\tilde{\beta} = \begin{cases} \beta^* & \text{on} \quad B^c_{\tilde{\delta}_1,\tilde{\delta}_2}. \\ (\beta_1^* + \delta', \beta_2^*, \cdots, \beta_m^* - \delta', \cdots, \beta_M^*) & \text{on} \quad B_{\tilde{\delta}_1,\tilde{\delta}_2}. \end{cases}$$

with $0 < \delta' < \min\left\{ \tilde{\delta}_1, \frac{\delta}{\int u_m dP(\mathbf{u})} \right\}$. It is easy to check that,

$$\bar{u}_m(\tilde{\beta}) = \bar{u}_m(\beta^*) - \delta' \int_{B_{\tilde{\delta}_1,\tilde{\delta}_2}} u_m dP(\mathbf{u}) > \theta_m.$$

That is, the new policy still satisfies the constraint for agent m. Also, it does not alter the decisions of any other constrained user which implies the constraints of the other users are also intact. Further,

$$\begin{aligned} \sum_k \bar{u}_k(\tilde{\beta}) &= \sum_k \bar{u}_k(\beta^*) + \delta' \int_{B_{\tilde{\delta}_1,\tilde{\delta}_2}} (u_1 - u_m) dP(\mathbf{u}) \\ &\geq \sum_k \bar{u}_k(\beta^*) + \delta' \tilde{\delta}_2 P(B_{\tilde{\delta}_1,\tilde{\delta}_2}) \ > \sum_k \bar{u}_k(\beta^*) \end{aligned}$$

contradicting the optimality of β^*. ◇

Proof of Theorem 4 : Any scheduler $\beta = (\beta_1, \beta_2, \cdots, \beta_M)$ satisfies that $\beta_m \geq 0$ and $\sum_m \beta_m = 1$ for almost all values of $\mathbf{u} = (u_1, u_2, \cdots, u_M)$. And so,

$$\Upsilon(\beta) = \int \sum_{2 \leq m \leq M} (u_m - u_1) \beta_m(\mathbf{u}) dP(\mathbf{u}) + \int u_1 dP(\mathbf{u}). \quad (13)$$

And the constraints are given by:

$$\int u_m \beta_m(\mathbf{u}) dP(\mathbf{u}) \geq \theta_m \text{ for } m \geq 2.$$

[2]Clearly $A = \cup_{n,l} B_{\frac{1}{n},\frac{1}{l}}$ and so, if such a $\tilde{\delta}_1, \tilde{\delta}_2$ does not exist then

$$P(A) = P\left(\cup_{n,m} \left\{ \beta_m^*(\mathbf{u}) > \frac{1}{n} \text{ with } \max_k u_k = u_1 \text{ and } (u_1 - u_m) > \frac{1}{l} \right\} \right) = 0.$$

The last term in equation (13) is independent of the scheduler β and can be omitted for optimization and further considering the dual variables $\varrho := (0, \varrho_2, \varrho_3, \cdots, \varrho_M)$ (note there is no constraint on the first agent and hence $\varrho_1 = 0$) we are interested in optimizing the following w.r.t. β:

$$\begin{aligned} \mathcal{L}(\beta; \varrho) &= \int \sum_{2 \leq m \leq M} [(u_m - u_1) \beta_m(\mathbf{u})] dP(\mathbf{u}) \\ &\quad + \sum_{2 \leq m \leq M} \varrho_m \left(\int u_m \beta_m(\mathbf{u}) dP(\mathbf{u}) - \theta_m \right). \quad (14) \end{aligned}$$

Note here that for agents with $\theta_m = 0$, i.e., for ones without constraint, we set $\varrho_m = 0$ for ease of notations. For any fixed ϱ, we are equivalently interested in:

$$\max_{\beta \in \Omega} \int \sum_{2 \leq m \leq M} [u_m(1 + \varrho_m) - u_1] \beta_m(\mathbf{u}) dP(\mathbf{u}). \quad (15)$$

Clearly the $\beta^*_{;\varrho} = (\beta^*_{1;\varrho}, \cdots, \beta^*_{M;\varrho})$ defined as below:

$$\begin{aligned} \beta^*_{1;\varrho}(\mathbf{u}) &= \Pi_{m \geq 2} \mathbb{1}_{\{[(1+\varrho_m)u_m - u_1] < 0\}} \text{ and} \\ \beta^*_{m;\varrho}(\mathbf{u}) &= \mathbb{1}_{\{\arg\max_{k \geq 2}[(1+\varrho_k)u_k - u_1] = m\}} (1 - \beta^*_{1;\varrho}(\mathbf{u})), \\ &\qquad \text{for any } m \geq 2 \end{aligned}$$

optimizes (15). It is easy see to that the above is same as the following ($\mathbf{1} := (1, \cdots, 1)$):

$$\beta^*_{m;\varrho}(\mathbf{u}) = \beta_m(\mathbf{1} + \varrho)(\mathbf{u}) \quad (16)$$

with $\beta_m(\mathbf{b})(\mathbf{u})$ defined as in (6), where $\mathbf{b} = \mathbf{1} + \varrho$. By Lemma 1, under **A**.4 this is the unique maximizer for any given ϱ.

The domain Ω is a convex closed subset of the Banach space, L^1. The objective function and the constraints are both linear. The constraints are functions from Ω to Euclidean space $\mathcal{R}^{n_c}$, where n_c represents the total number of constraints. We consider normed space $\mathcal{R}^{n_c}$ with usual partial order. By Theorem 2, which establishes the existence of optimizer and **A**.3, the Duality theorem on normed linear spaces, [7, Theorem 5] is applicable (see also [8]). Further, the optimizer given by (16) is unique for each tuple $\varrho = \{\varrho_m\}$ and these two give us the following:

- By the first statement of [7, Theorem 5]

$$\max_{\substack{\beta \in \Omega; \\ \bar{u}_m(\beta) \geq \theta_m, \forall m}} \Upsilon(\beta) = \min_{\varrho \geq 0} \max_{\beta} \mathcal{L}(\beta, \ \varrho)$$

  and there exists a ϱ^* which satisfies the constraints and which achieves the minimum on the right hand side.
- By Theorem 2, there exists an optimizer β^* of (5). Then, by the second statement of [7, Theorem 5] β^* also maximizes $\mathcal{L}(\beta, \varrho^*)$. By Lemma 1 there exists an unique maximizer $\beta^*_{;\varrho^*}$ for $\mathcal{L}(\beta, \varrho^*)$ and hence (see (16)):

$$\beta^* = \beta^*_{;\varrho^*} = \beta(\mathbf{1} + \varrho^*) \quad (17)$$

  Note here $\varrho^*_m = 0$ for $\theta_m = 0$.

By Theorem 3 any maximizer satisfies the constraints and by Theorems 5 and 3 there is a unique maximizer in Ω'. ◇

Remarks about discrete utility case: Equation (17) in the proof of Theorem 4 is the crucial step to show that an element from finite dimensional Ω' maximizes the constrained optimization (5). This step is true by Lemma 1, which establishes the existence of an unique optimizer for the Lagrangian objective function $\mathcal{L}(\beta, \varrho)$ given by (14), with any arbitrary Lagrange multiplier ϱ. The lemma is true only under **A**.4, i.e., when the instantaneous utilities are continuous random variables. One can easily construct examples with discrete utilities when uniqueness fails and we present one such a simple example in Appendix B. But the rest of the arguments would be same for discrete rates and in this case the optimizer of (5) still maximizes $\mathcal{L}(\beta, \varrho^*)$, however it can also be a convex combination of all the maximizers of $\mathcal{L}(\beta, \varrho^*)$ and hence may not be an element from Ω'.

Lemma 1. *Under* **A***.4, for any given ϱ the scheduler given by (16) is the unique maximizer of (14).*

Proof: By adding the constant term $\int u_1 dP(u)$ to (14) and noticing that $\beta_1 = 1 - \sum_{m \geq 1} \beta_m$ we are equivalently interested in optimizing (with $\varrho_1 = 0$):

$$\tilde{\mathcal{L}}(\beta) := \int \sum_m u_m(1+\varrho_m)\beta_m(\mathbf{u})dP(u).$$

The maximizer $\beta^*_{;\varrho}$ given by (16) is clearly the point-wise maximizer of the term inside the integration. Define the following random variables:

$$m^* = \arg\max_m u_m(1+\varrho_m), \; u^* = \max_m u_m(1+\varrho_m) \text{ and}$$
$$\tilde{u}^* = \max_{m \neq m^*} u_m(1+\varrho_m).$$

Under **A**.4, $P(u_k = u_m) = 0$ for any $k \neq m$ and hence

$$\begin{aligned} P(u^* = \tilde{u}^*) &= \sum_k P(\max_m u_m = \tilde{u}^*;\, m^* = k) \\ &= \sum_k P(u_k = \max_{m \neq k} u_m;\, m^* = k) \\ &\leq \sum_k \sum_{m \neq k} P(u_k = u_m;\, m^* = k) \\ &\leq \sum_k \sum_{m \neq k} P(u_k = u_m) = 0. \end{aligned} \tag{18}$$

If there exists any other maximizer $\tilde{\beta}$ of (14) and say

$$P(A) > 0 \text{ with } A := \{\mathbf{u} : \beta^*_{;\varrho}(\mathbf{u}) \neq \tilde{\beta}(\mathbf{u})\}.$$

Because u^* is a point-wise maximum value and $\tilde{u}^*$ is the second maximum,

$$\begin{aligned} A &= \cup_k B_k \cup \left\{\mathbf{u} : \beta^*_{;\varrho}(\mathbf{u}) \neq \tilde{\beta}(\mathbf{u}),\; u^* = \tilde{u}^*\right\} \text{ with} \\ B_k &:= \left\{\mathbf{u} : \beta^*_{;\varrho}(\mathbf{u}) \neq \tilde{\beta}(\mathbf{u});\; u^* - \tilde{u}^* > \frac{1}{k}\right\} \end{aligned}$$

In view of (18) and because $P(A) > 0$, there exists at least one $\tilde{k}$ such that:

$$P(B_{\tilde{k}}) > 0.$$

Because $\sum_m \tilde{\beta}_m = 1$ for all $\mathbf{u}$,

$$\begin{aligned} &\tilde{\mathcal{L}}(\beta^*_{;\varrho}) - \tilde{\mathcal{L}}(\tilde{\beta})) \\ &= \int u^* dP(\mathbf{u}) - \int \sum_m (u_m(1+\varrho_m)\tilde{\beta}_m(\mathbf{u}))dP(\mathbf{u}) \\ &= \int u^* \sum_m \tilde{\beta}_m(\mathbf{u})dP(\mathbf{u}) \\ &\quad - \int \sum_m (u_m(1+\varrho_m)\tilde{\beta}_m(\mathbf{u}))dP(\mathbf{u}) \\ &= \int \sum_m \tilde{\beta}_m(\mathbf{u})(u^* - u_m(1+\varrho_m))dP(\mathbf{u}) \\ &= \int_A \sum_m \tilde{\beta}_m(\mathbf{u})(u^* - u_m(1+\varrho_m))dP(\mathbf{u}) \\ &\geq \int_A \sum_m \tilde{\beta}_m(u^* - \tilde{u}^*)dP(\mathbf{u}) = \int_A (u^* - \tilde{u}^*)dP(\mathbf{u}) \\ &\geq \int_{B_{\tilde{k}}} (u^* - \tilde{u}^*)dP(\mathbf{u}) \geq \int_{B_{\tilde{k}}} \frac{1}{\tilde{k}}dP(\mathbf{u}) = \frac{1}{\tilde{k}}P(B_{\tilde{k}}) > 0. \end{aligned}$$

which is a contradiction to the optimality of $\tilde{\beta}$. ◇

Theorem 5. *There exists at maximum one element in Ω' which satisfies all the constraints of (5).*

Proof: Consider the case with $M-1$ constraints. The general case goes through in a similar way. If possible, consider $\mathbf{b} \neq \mathbf{c}$, such that both of them satisfy the constraint equations. Recall by Theorem 4 that $c_1 = b_1 = 1$. Define the following sets:

$$A_{j,k}(\mathbf{b}) := \{u_j b_j < u_k b_k\} \text{ for all } 1 \leq j,k \leq n \text{ with } j \neq k.$$

Define, $A_m(\mathbf{b}) := \cap_{j \neq m} A_{j,m}(\mathbf{b})$. Clearly,

$$\bar{u}_m(\mathbf{b}) = E[u_m \beta_m(\mathbf{b})(\mathbf{u})] = E[u_m \mathbb{1}_{A_m(\mathbf{b})}].$$

W.l.g. say $c_M > b_M$. Then clearly,

$$A_{1,M}(\mathbf{b}) = \{u_1 < u_M b_M\} \subset \{u_1 < u_M c_M\} = A_{1,M}(\mathbf{c}).$$

Since agent M's constraint is satisfied at both $\mathbf{b}$ and $\mathbf{c}$ (i.e., $\bar{u}_M(\mathbf{b}) = \theta_M = \bar{u}_M(\mathbf{c})$), the above relation implies there exists at least one m_1 with $1 < m_1 < M$ and such that

$$A_{m_1,M}(\mathbf{b}) \supset A_{m_1,M}(\mathbf{c}) \Rightarrow \frac{b_M}{b_{m_1}} > \frac{c_M}{c_{m_1}}.$$

W.l.g. let $m_1 = M-1$. As,

$$\frac{b_M}{b_{M-1}} > \frac{c_M}{c_{M-1}} \text{ and } c_M > b_M \text{ we have } c_{M-1} > b_{M-1}.$$

These in turn imply (for agent $M-1$) that:

$$A_{1,M-1}(\mathbf{b}) \subset A_{1,M-1}(\mathbf{c}) \text{ and } A_{M,M-1}(\mathbf{b}) \subset A_{M,M-1}(\mathbf{c}).$$

Again because $\bar{u}_{M-1}(\mathbf{b}) = \theta_{M-1} = \bar{u}_{M-1}(\mathbf{c})$, there should be at least one m_2 with $1 < m_2 < M-1$ and such that

$$A_{m_2,M-1}(\mathbf{b}) \supset A_{m_2,M-1}(\mathbf{c}) \text{ which implies } \frac{b_{M-1}}{b_{m_2}} > \frac{c_{M-1}}{c_{m_2}}.$$

W.l.g. let $m_2 = M-2$ and this implies $c_{M-2} > b_{M-2}$,

$$\begin{aligned} \frac{b_{M-1}}{b_{M-2}} &> \frac{c_{M-1}}{c_{M-2}} \text{ and} \\ \frac{b_M}{b_{M-2}} &= \frac{b_M}{b_{M-1}}\frac{b_{M-1}}{b_{M-2}} > \frac{c_M}{c_{M-1}}\frac{c_{M-1}}{c_{M-2}} = \frac{c_M}{c_{M-2}}. \end{aligned}$$

Thus we have (now for agent $M-2$)

$$A_{1,M-2}(\mathbf{b}) \subset A_{1,M-2}(\mathbf{c}), \;\; A_{M,M-2}(\mathbf{b}) \subset A_{M,M-2}(\mathbf{c})$$

and $A_{M-1,M-2}(\mathbf{b}) \subset A_{M-1,M-2}(\mathbf{c})$.

Thus and because $\bar{u}_{M-2}(\mathbf{b}) = \bar{u}_{M-2}(\mathbf{c})$ there exists an $m_3 = M-3$ (w.l.g.) such that

$$A_{M-3,M-2}(\mathbf{b}) \supset A_{M-3,M-2}(\mathbf{c}) \;\Rightarrow\; \frac{b_{M-3}}{b_{M-2}} > \frac{b_{M-3}}{c_{M-2}}.$$

Continue in the same way up to 3 and for $\bar{u}_3(\mathbf{b}) = \bar{u}_3(\mathbf{c})$ we will need that $c_2 > b_2$ so that $b_3/b_2 > c_3/c_2$. This means (calculating as before) for agent 2 that:

$$A_{1,2}(\mathbf{b}) \subset A_{1,2}(\mathbf{c}), \text{ and}$$

$$A_{m,2}(\mathbf{b}) \subset A_{m,2}(\mathbf{c}) \text{ for all } m.$$

This implies that $\bar{u}_2(\mathbf{b}) < \bar{u}_2(\mathbf{c})$, which contradicts the fact that both of them satisfy all the constraints. ◇

Appendix B: A discrete utility example

We consider a simple case with two agents. The first agent has $u_1 \in \{20, 10\}$ and $Prob(u_1 = 10) = Prob(u_1 = 20)$ while for the second agent $Prob(u_2 = 5) = 1$. In this case the scheduler $\beta = (\beta_1, \beta_2)$, with β_1 representing the probability with which agent is scheduled when its utility is 20 while β_2 represents the same when agent 1's utility is 10. In this case,

$$\begin{aligned} \bar{u}_1(\beta) &= 20 * 0.5 * \beta_1 + 10 * 0.5 * \beta_2 \text{ and} \\ \bar{u}_2(\beta) &= 5 * (0.5 * (1-\beta_1) + 0.5 * (1-\beta_2)). \end{aligned}$$

For example, if the fairness constraint is $\bar{u}_2 \geq \theta$, it can be satisfied by the following two β's:

$$(1, 1-\theta/2.5) \text{ or } (1-\theta/2.5, 1).$$

Thus uniqueness of zero of 'fairness constraints equation' is not true in discrete case. In view of Theorem 3, it is easy to see that among the above two, the optimizer of 5 is $(1, 1-\theta/2.5)$. This scheduler obtained by changing the $\beta^{eff} = (1,1)$ (efficient scheduler) only in the second component. This component corresponds to $u_1 = 10$ and hence is the one with lesser $u_1 - u_2$.

Thus, while choosing the 'optimal' zero of fairness constraint, one first needs to change the beta's (from beta's corresponding to efficient scheduler) corresponding to lowest difference utilities and then to the next lower utilities and so on. That is exactly the intuitive reason why priority type scheduler works with continuous utilities. We plan to explore further these ideas in future.