Price Based Protocols For Fair Resource Allocation: Convergence Time Analysis and Extension to Leontief Utilities*

Ashish Goel † Stanford University

Hamid Nazerzadeh ‡ Microsoft Research

Abstract

We analyze several distributed, continuous time protocols for a fair allocation of bandwidths to flows in a network (or resources to agents). Our protocols converge to an allocation which is a logarithmic approximation, simultaneously, to all canonical social welfare functions (i.e. functions which are symmetric, concave, and non-decreasing). These protocols can be started in an arbitrary state. While a similar protocol was known before, it only applied to the simple bandwidth allocation problem, and its stability and convergence time was not understood. In contrast, our protocols also apply to the more general case of Leontief utilities, where each user may place a different requirement on each resource. Further, we prove that our protocols converge in polynomial time. The best convergence time we prove is $O\left(n \log \frac{n c_{\max} a_{\max}}{c_{\min} a_{\min}}\right)$, where n is the number of agents in the network, $c_{\max}$ and $c_{\min}$ are the maximum and minimum capacity of the links, and a_{max}, a_{min} are the largest and smallest Leontief coefficients, respectively. This time is achieved by a simple MIMD (multiplicative increase, multiplicative decrease) protocol which had not been studied before in this setting. We also identify combinatorial properties of these protocols that may be useful in proving stronger convergence bounds. The final allocations by our protocols are supported by usage-sensitive dual prices which are fair in the sense that they shield light users of a resource from the impact of heavy users. Thus our protocols can also be thought of as efficient distributed schemes for computing fair prices.

1 Introduction

In many problems where a centralized planner needs to distribute a set of resources among a group of individuals, the individual utility functions are known but there is no well defined notion of how to combine individual utilities into a single aggregate utility function(i.e. a social welfare function). For instance, in most real-life settings, increasing efficiency (i.e. the sum of individual utilities) is an important goal. At the same time, achieving fairness in the allocations to different individuals is also desirable. It has recently been shown that it is possible to simultaneously approximate a large class of social welfare functions [22, 17, 23, 4]. Efficient distributed algorithms have been presented for cases where the individual utility functions can be modeled as the amount of flow allocated to that individual in a multi-commodity like problem, and price based protocols have been presented when the flow for an individual in this multi-commodity problem must use a single route fixed in advance [10].

In this paper, we extend the efficient distributed algorithm mentioned above to the case where the individuals have Leontief utilities. The distributed algorithm must be initialized to a special state and then proceed in a synchronized fashion. We then present two simple price based protocols which achieve the same equilibrium as the distributed algorithm. These protocols can start with an arbitrary allocation, are stable around the equilibrium point, and converge quickly. The stability and convergence of price based protocols was left as an open problem even for the simple case of multi-commodity flows. Our price based protocols are simple and efficient enough to be implemented in a TCP-like fashion in computer networks. We believe

*An earlier version of this work has appeared in the proceedings of SODA'08.

†Department of Management Science and Engineering, and (by courtesy) Computer Science, Stanford University, CA. Research supported in part by NSF ITR grant 0428868, an NSF CAREER award. Email: ashishg@stanford.edu

‡Microsoft Research, Cambridge, MA. Email: hamidnz@microsoft.com

that our results will also be interesting in designing exchange markets which lead to an approximately fair outcome, under all reasonable definitions of fairness.

In order to describe our results in more detail and put them in context, we will first need to define and motivate the problem and also briefly survey recent related work.

Problem description

The problem of allocation of m resources among n agents with Leontief utilities is described by two positive matrices, denoted C and A. Matrix $C = (c_1, \cdots, c_m)$ indicates the amount of available resources. Let Y be an allocation of resources among the agents, i.e. y_{ij} unit of resource j is allocated to agent i. Let $S_i = \{j | a_{ij} > 0\}$. The utility of agent i from allocation Y is:

$$u_i(y_i) = \min_{j \in S_i} \{\frac{y_{ij}}{a_{ij}}\}.$$

In other words, S_i is the set of resources that agent i uses and a_{ij} is the rate of the consumption of each unit of resource j. It is convenient to characterize an allocation by a vector $\mathbf{x} = (x_1, \cdots, x_n)$, and set $y_{ij} = a_{ij} x_i$. The resource allocation problem with Leontief utilities generalizes the problem of *bandwidth allocation* in networks when the route used by each connection is fixed in advance, i.e. $a_{ij} \in \{0, 1\}$.

A resource allocation $\mathbf{x}$ is *feasible* if it respects capacity constraints, i.e. $\forall j : \sum_{i: j \in S_i} x_i a_{ij} \leq c_j$. In the sequel, without loss of generality, we assume $\max\{a_{ij}\} = 1$ and $\min\{c_i\} = 1$. Also, let $a_{\min} = \min_{a_{ij} > 0}\{a_{ij}\}$ and $c_{\max} = \max_i\{c_i\}$.

We are interested in finding a feasible allocation which simultaneously optimizes a large class of objective functions. Let U be an n-variate real-valued function. We say that U is a *canonical utility function* if U is symmetric in its arguments, concave, non-decreasing, and $U(0) = 0$. Let U^* denote the maximum value of $U(\mathbf{x})$ subject to feasibility of $\mathbf{x}$.

Definition 1. *Feasible allocation* $\mathbf{x}$ *is* α*-fair if*

$$\alpha \geq \sup_{U: U \text{ is canonical}} \frac{U^*}{U(\mathbf{x})}.$$

Let $\rho = \max\left\{n, m, c_{\max}, \frac{1}{a_{\min}}\right\}$. We show there always exists a feasible solution $\mathbf{x}$ which is $O(\log \rho)$-fair. The class of utility functions we consider is not arbitrarily chosen; a detailed motivation and history is in [10]. This is a large class, and contains the important subclass $\sum_i f(x_i)$ where f is a uni-variate concave function (f must also be non-decreasing and $f(0)$ must be 0). Concavity is a natural restriction, since it corresponds to the “law of diminishing returns” from economics. Symmetry corresponds to saying that all users are equally important[1]. The requirement for U being non-decreasing is natural for a resource allocation problem. The requirement that $U(0) = 0$ is also natural in many (but not all) settings. The class of canonical utility functions includes important special functions such as min, $\sum_i x_i$, $\sum_i \log(1 + x_i)$, and $P_j(x)$ = the sum of the j smallest x_i's. The class of canonical utility functions also contains a series of functions which together capture max-min fairness. Most interestingly, there is a concrete connection between this class and our intuitive notion of fairness. Suppose there exists some function which measures the fairness of an allocation. It seems natural that the allocation (x_1, x_2) should be deemed as fair as (x_2, x_1) and less fair than $\left(\frac{x_1+x_2}{2}, \frac{x_1+x_2}{2}\right)$. This assumption implies that for any natural definition of fairness, maximizing fairness should be equivalent to maximizing some symmetric concave function; certainly, all the definitions of fairness that we found in literature are of this form.

Some important utility functions satisfy symmetry, concavity, and the non-decreasing property but are not 0 at 0. One example is the function $\sum_i \log x_i$, which is particularly important in multiple contexts [24, 12].

[1] Notice that the constraints are not required to be symmetric, and hence, the optimum solution need not be symmetric even though the objective function is symmetric.

All our results hold for this more general class of functions as well in the resource augmented setting; details are similar to those in [10] and are omitted.

As detailed later, centralized algorithms for solving this problem were already known. Our focus in this paper is on designing simple protocols supported by a pricing scheme and analyzing their convergence and fairness ratio[2].

Price based protocols: In order to implement these ideas in the setting of a centralized social planner, it would also be important to derive a natural pricing mechanism which implements simultaneous optimization. We define a price based protocol as one where:

1. Each resource j maintains a dual price (or "shadow" price) l_j^i for each agent i. The price is allowed to depend only on the vectors $\langle x_i : j \in S_i \rangle$, $\langle a_{ij} : j \in S_i \rangle$, and capacity of the resource, i.e. only on information which is locally available to the resource.

2. Each agent computes an aggregate dual price $w_i = \sum_{j \in S_i} a_{ij} l_j^i$.

3. The evolution of the utility for agent i is a function of $x_i(t)$ and $w_i(t)$:

$$\frac{dx_i}{dt} = f_i\left(x_i(t), w_i(t)\right).$$

In addition, we assume the initial value of x_i is at most $c_{\max}$. Informally, each agent has an endowment that he uses to buy resources at the prevailing dual prices. This general approach has been used before in many settings. [26, 3, 15, 21, 14]. However, here the prices l_j^i are allowed to depend on the entire vector of utilities observed at a resource rather than just the sum. This is a key difference, and allows us to approximate all canonical functions simultaneously. Our goal in this paper is to derive price based protocols that are (a) as general as possible, (b) converge quickly to a unique equilibrium, and (c) guarantee a good fairness ratio.

Along these lines, Cho and Goel [10] gave a simple algorithm, called Distributed-Majorized, for obtaining fair bandwidth allocation given fixed routes: In this algorithm, at time zero, all the flows are initialized to zero and all the dual price of edges initialized to have a small dual price, denoted l_0. Agent i increases its flow as long as its dual price is less than 1. Let $\Lambda_e(t)$ be the congestion of edge e at time t, i.e. $\Lambda_e(t) = \sum_{i: e \in S_i} \frac{x_i(t)}{c_e}$. The dual price $l_e(t)$ computed for edge e, is equal to $l_0 e^{\delta \Lambda_e(t)}$, where l_0 is a constant and $\delta = O(\log \rho)$. Surprisingly, this simple primal dual scheme gives an $O(\log \rho)$-fair allocation which is almost optimum. They also obtained a price based protocol for their restricted setting (i.e. all $a_{ij} \in \{0, 1\}$) but its convergence time and stability properties were left as open questions.

Our Results

We extend the work of Cho and Goel [10] in two significant ways.

1. We extend their distributed algorithm (as well as their price based protocol) to Leontief utilities (they assumed $a_{ij} \in \{0, 1\}$). In particular, this allows us to handle the important sub-case of weighted utilites, i.e. where the utility of agent i needs to be weighted by some factor b_i. In communication networks, these weights could represent "Quality of Service" parameters. Our distributed algorithm (and our price based protocols) find an $O(\log \rho)$-fair resource allocation.

2. We do a convergence analysis of our price based protocols, and show that they converge very fast (i.e. in polynomial time). No convergence time bounds were known before even for the case $a_{ij} \in \{0, 1\}$.

[2] The main difference between an algorithm and a protocol is that algorithms start from precise initial conditions.

More specifically, we define a broad class of price based protocols, that we call truncated protocols (section 2). We prove that all truncated protocols achieve a fairness ratio of $O(\log\rho)$ and have a unique equilibrium. We then show a decomposition result that is useful in analyzing this class of protocols: the convergence time of a truncated protocol is at most $O(nt_{max})$ where t_{max} is the time required for a single agent's allocation to stabilize assuming that all other allocations are held fixed. We then present two specific truncated protocols. One of them (section 3) is almost MIMD (Multiplicative Increase Multiplicative Decrease) protocol and we prove that the convergence time of this protocol is $O\left(\frac{1}{\gamma}n\log\rho\right)$, where γ is a rate parameter. In the other protocol (section 4), the rate of change of x_i is almost proportional to $\log w_i$ and this converges in $O(nc_{\max})$ time. We also analyze a variation of this protocol which converges in $O(nr)$ time, where r is the maximum number of the agents share a single resource. Finally, we show that these protocols have much faster convergence over a single link network (section 5).

The extension to Leontief utilities extends the proof techniques of Cho and Goel [10]. The convergence time analysis is novel, and offers additional insight into price based protocols. For ease of exposition, we present the extension to Leontief utilities before the convergence time results. We present all our results in a continuous time setting; discretization into small steps is straightforward.

Related Work

Building on a series of papers about multi-objective fairness [22, 17], Kleinberg and Kumar [23] studied the problem of bandwidth allocation in a centralized setting with multiple fairness objectives. Goel and Meyerson [16] and Bhargava, Goel, and Meyerson [4] later expanded this work to a large class of linear programs and related it to simultaneous optimization [18]. In particular, the above sequence of papers resulted in a centralized bandwidth allocation algorithm for computing a single allocation which is simultaneously an $O(\log\rho)$ approximation for all canonical utility functions. Goel and Meyerson [16] build on the notion of *majorization* due to Hardy, Littlewood, and Polya [19, 20, 25] (see Appendix A). This leaves open the following question: *Can there be efficient distributed algorithms which achieve the same results?* Cho and Goel [9] made some partial progress toward this problem by giving a centralized algorithm which maintains only one set of dual costs. Later, Cho and Goel presented a simple distributed algorithm a simple protocol supported by prices which achieves the same result [10].

The problem of fair resource allocation where the sequence of jobs arrives in an online fashion is also interesting [18]. Recently, Buchbinder and Naor [5, 6] presented improved online algorithms for fair resource allocation problems, along with an interesting centralized primal-dual framework.

Another related line of work to ours is that of all-norm approximation algorithms. In the context of job scheduling, [7, 1, 2] provide algorithms that find an assignment of jobs to machines such that the solution simultaneously approximates the optimal assignment with respect to any norm $p > 1$.

Leontief utilities are an important case of CES (Constant Elasticity of Substitution) utilities, and have also been studied in the exchange market context [11, 27]. Recently, Codenotti et. al. [12] showed that Leontief economies encode nonzero Sum two-player games which implies computing the equilibrium prices for Leontief econimies is PPAD-complete [8, 13].

2 Truncated Protocols

We first present a centralized algorithm that we call the Distributed-Weighted-Majorization[3] algorithm which must start from precise initial conditions and which is $O(\log\rho)$-fair for the resource allocation problem with Leontief utilities. In a TCP-like protocol or an implementation in a dynamic market, the allocation should settle in an equilibrium even if it starts with arbitrary initial values. Accordingly, we then define a class of price based protocols that we call "truncated protocols" and show that such protocols converge to a unique

[3]The term majorization refers to the proof-technique used; since the proof is in the appendix, a definition of majorization has also been postponed to the appendix.

equilibrium which is the same as that of the centralized algorithm. We also prove that the convergence time of these protocols is at most $O\left(nt_{max}\right)$ where t_{max} is the maximum time required for a single agent to converge into the equilibrium, given that all other agents are held fixed.

2.1 The Centralized Algorithm

Define $\mu = \rho^3$ and $\eta = \log_\mu \frac{1}{a_{\min}} + 1$. Note that that $\eta \leq \frac{4}{3}$; also, for the bandwidth allocation problem, $\eta = 1$. We also define γ as a rate parameter which controls the rate of the changes of the utilities.

The algorithm below computes fair allocation and a corresponding dual prices. It is similar to the Distributed-Majorization algorithm [10]. However, the dual prices are computed differently.

Distributed-Weighted-Majorization Algorithm:

The congestion of resource j is defined as $\Lambda_j\left(t\right) = \left(\sum_{i:j\in S_i} a_{ij}x_i\left(t\right)\right)/c_j$. Let $l_j = \mu^{\eta\Lambda_j - 1}$ and $w_i = \sum_{j\in S_i} a_{ij}l_j$. At time 0, for $1 \leq i \leq n$, initialize $x_i = \frac{\eta}{\rho}$. Update the amount of x_i according to the differential equation below:

$$\frac{dx_i}{dt} = \begin{cases}\gamma, & w_i < 1\\ 0, & w_i \geq 1.\end{cases}$$

Lemma 1 (Feasibility)**.** *The allocation found by the Distributed-Weighted-Majorization algorithm is feasible.*

Proof. Consider resource j and let i be one of the agents in $\mathbf{argmax}_{i:j\in S_i} x_i$. Let t be the first time $w_i\left(t\right)$ becomes equal to 1. At time t:

$$a_{\min}\mu^{\eta\Lambda_j(t)-1} \leq a_{ij}\mu^{\eta\Lambda_j(t)-1} \leq \sum_{k:k\in S_i} a_{ik}l_k\left(t\right) = w_i = 1.$$

Note that by definition of x_i, Λ_j remains constant after time t. Therefore, $\eta\Lambda_j \leq \log_\mu \frac{1}{a_{\min}} + 1 = \eta$ which leads to the lemma.

□

Along the lines of [10, 15], the discrete implementation of the Distributed-Weighted-Majorization algorithm is also possible. The discrete jumps of update rules should be chosen small enough to maintain feasibility.

Theorem 2. *The allocation found by the Distributed-Weighted-Majorization algorithm is $O\left(\log\rho\right)$-fair.*

The detailed proof is given in appendix B.

2.2 Propertiess

Intuitively, in a fair allocation, an agent with a small utility should not be penalized if the other agents are using large amount of the shared resources. Suppose $\mathbf{x} = (x_1, \ldots, x_n)$ is the resource allocation at time t. We define Λ_j^i as the congestion of resource j, from the perspective of agent i, and it is computed by truncating the allocation of other agents which shares resource j:

$$\Lambda_j^i\left(t\right) = \left(\sum_{k:j\in S_k} a_{kj}\min\{x_i\left(t\right), x_k\left(t\right)\}\right)/c_j \tag{1}$$

Let $\mu = \rho^3$; the truncated dual price of resource j for agent i at time t, denoted $l_j^i(t)$, is defined as:

$$l_j^i(t) = \mu^{\eta \Lambda_j^i(t) - 1}. \tag{2}$$

Consequently, the *truncated total dual price* for agent i at time t is equal to $\sum_{j \in S_i} a_{ij} l_j^i(t)$. In the sequel, we use $w_i(t)$ to refer to the truncated dual price. Also, we omit the parameter t when it is clear from the context. Before going further, we need some additional definitions:

Definition 2 (Monotonicity)**.** *A price based protocol $\mathcal{P}$ is monotone if for any two allocations $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$, $x_i = y_i$, and $x_j \leq y_j$, for $i \neq j$, implies $\frac{dx_i}{dt} \geq \frac{dy_i}{dt}$.*

Definition 3 (Truncated Protocol)**.** *A price based protocol $\mathcal{P}$ is* truncated *if it has the following properties:*

1. *If w_i, the truncated dual price for agent i, is less than 1, then $\frac{dx_i}{dt} > 0$. If $w_i > 1$, then $\frac{dx_i}{dt} < 0$. And if $w_i = 1$ then $\frac{dx_i}{dt} = 0$.*
2. *It is monotone.*

To illustrate the concept, consider truncated protocol $\mathcal{P}$ and two allocations $\mathbf{x}$ and $\mathbf{y}$ satisfying the assumptions in Definition 2. Also, let w_i and w_i' be the corresponding dual prices for these allocations. By assumptions (1) and (2) in Definition 3, if $w_i' > 1$, then y_i decreases at a faster rate than x_i, note that x_i might be even increasing in this case. Also, if $w_i < 1$, the rate of the increase of x_i is more than the rate of y_i.

2.3 Convergence time analysis

In this section we analyze the convergence of truncated protocols. In the sequel, we call the utility of an agent i *stable* if $w_i = 1$.

Definition 4 ($\mathbf{t}_{\max}(\mathcal{P})$)**.** *For a truncated protocol $\mathcal{P}$, $t_{\max}(\mathcal{P})$ is defined as the maximum time it would take for the utility of an arbitrarily chosen agent to become stable, assuming all the other utilities in the network are held fixed.*

Recall that we assume the initial amount of x_i, $1 \leq i \leq n$, is at most $c_{\max}$. When it is clear from the context, we omit the parameter $\mathcal{P}$ in the definition above. The abstractions above enable us to prove strong fairness and convergence properties for truncated protocols. While a similar protocol was presented by Cho and Goel for the restricted case of bandwidth allocation, no convergence time results were known even in this simpler setting.

Theorem 3. *Consider a truncated protocol $\mathcal{P}$:*

1. *It always converges to an equilibrium which is unique for all the truncated protocols.*
2. *The resource allocation at the equilibrium is $O(\log \rho)$-fair.*
3. *$\mathcal{P}$ takes at most $O(nt_{\max})$ time to converge.*

Proof. Let $\mathbf{f} = (f_1, \cdots, f_n)$ be the allocation found by the Distributed-Weighted-Majorized algorithm, starting with the precise initial conditions. Without loss of generality assume $f_1 \leq f_2 \leq \ldots \leq f_n$. Consider truncated protocol $\mathcal{P}$. Let x_k be the current utility of agent k, $1 \leq k \leq n$. We define $2n$ phases, each of which starts right after the previous one. For $1 \leq i \leq n$:

(a) Phase $2i - 1$ ends when for $i \leq k \leq n$, $x_k \geq f_i$.

(b) Phase $2i$ ends when $x_i = f_i$.

Of course, it is not immediately clear that the protocol will ever reach the end of phase $2n$. We will show, by induction, that the protocol will converge. Moreover, we claim the following properties hold:

(c) For $i \leq k$, after end of phase $2i-1$, x_k never drops below f_i.

(d) After phase $2i$ to the end, x_i remains equal to f_i.

(e) Each phase takes at most $t_{\max}$ time.

Note if (a) to (e) hold, after phase $2n$, $\mathbf{x} = \mathbf{f}$; and all utilities are stable. Therefore, the unique equilibrium is $\mathbf{f}$ which by Theorem 2 is $O(\log \rho)$-fair. Also, convergence takes at most $2nt_{\max}$ time.

We prove these by induction on the phases. Define $f_0 = 0$, so we can assume that induction basis holds. Consider the protocol at the beginning of phase $2i-1$, $i \geq 1$. Suppose the utility of an agent $k \geq i$ is less than f_i. First observe that for resource $j \in S_k$:

$$\begin{aligned}
\Lambda_j^k &= \left(\sum_{l:j\in S_l} a_{lj} \min\{x_k, x_l\}\right)/c_e \\
&= \left(\sum_{l:j\in S_l, l<i} a_{lj} \min\{x_k, x_l\}\right)/c_e + \left(\sum_{l:j\in S_l, l\geq i} a_{lj} \min\{x_k, x_l\}\right)/c_e \\
&\leq \left(\sum_{l:j\in S_l, l<i} a_{lj} f_l\right)/c_e + \left(\sum_{l:j\in S_l, l\geq i} a_{lj} x_k\right)/c_e \qquad (3)
\end{aligned}$$

Inequality (3) follows by the induction hypothesis which implies properties (c) and (d) for phase $2i-2$. Let $\bar{\Lambda}_j^k$ be the congestion of resource j in the Weighted-Distributed-Majorization algorithm at the time t such that $x_k(t) = f_i$, i.e.

$$\bar{\Lambda}_j^k = \left(\sum_{l:j\in S_l, l<i} a_{lj} f_l\right)/c_e + \left(\sum_{l:j\in S_l, l\geq i} a_{lj} f_i\right)/c_e$$

Therefore, as long as $x_k < f_i$, by inequality 3, $\Lambda_j^k \leq \bar{\Lambda}_j^k$. Hence, w_k is less than 1 and $\frac{dx_k}{dt}$ is positive. Thus, phase $2i-1$ ends in finite time. A similar argument shows that after this phase, for $k \geq i$, $\frac{dx_k}{dt}$ is nonnegative at f_i. Thus, x_k never drops below f_i; this gives us the property (c).

After this phase, similarly, as long as $x_i > f_i$, $\frac{dx_i}{dt}$ is negative. Hence, phase $2i$ also ends in finite time. Note that because of property (c) and the truncated dual pricing, Λ_j^i remains constant after phase $2i$. Hence, by the induction hypothesis, property (d) holds after phase $2i$.

Now, using monotonicity, we show that phase $2i-1$ takes at most $t_{\max}$. Consider agent $k \geq i$ with utility less than f_i; if there is no such agent we are done. By condition (c), for $l < i$, x_l is fixed at f_l. Consider the scenario in which for $l \neq k$ and $l \geq i$, x_l is held fixed at f_i. Let t' be the time it takes for x_k to become stable, by definition, $t' < t_{\max}$. We have already shown that the stable point of x_k is at least f_i. Therefore, in this scenario x_k to reach f_i in at most t' time. Hence, by monotonicity, and because the prices are truncated, the time it takes for x_k, in phase $2i-1$, to reach f_i is at most t'. Similarly, for phase $2i$, define t' as the time it takes for i to become stable assuming x_l, for $l > i$, is held fixed at f_i and $x_l = f_l$, for $l < i$. Again $t' \leq t_{\max}$. This proves the induction step for property (e).

Therefore, the protocol reaches end of phase $2n$ and properties (a)-(e) hold afterward. Hence, the stable point of the protocol is $\mathbf{x} = \mathbf{f}$ which is $O(\log \rho)$-fair. Also, convergence takes at most $2nt_{\max}$ time. □

In the next two sections we define and analyze two truncated protocols.

3 A Primal Protocol

We define a protocol in which the rate of change of a utility is (mostly) proportional to its current value. We call this protocol the *primal protocol*. We need one more definition to describe this protocol. Function σ is defined as:

$$\sigma(z) = \begin{cases} 1, & z < 1 \\ 0, & z = 1 \\ -1, & z > 1 \end{cases}$$

In a primal protocol with the rate parameter γ, denoted $\mathcal{P}_\gamma$, agent i, $1 \le i \le n$, updates its utility according to the differential equation:

$$\frac{dx_i}{dt} = \begin{cases} \frac{\gamma}{2n}, & x_i \le \frac{1}{2n} \\ \sigma(w_i)\gamma x_i, & o.w. \end{cases} \tag{4}$$

Proposition 4. *Primal protocols are truncated .*

Proof. With the definition of σ, to prove that primal protocols satisfies the property (1) in Definition 3, we only need to show that if $x_i < \frac{1}{2n}$, then $w_i < 1$. Recall that $w_i = \sum_{j \in S_i} \mu^{\eta \Lambda_j^i - 1}$. Because Λ_j^i is the truncated congestion, $\Lambda_j^i < n \times \frac{1}{2n} = \frac{1}{2}$. Therefore, $w_i < m\mu^{\frac{\eta}{2}-1} \le \rho^{\frac{1}{3}+\frac{4}{6}-1} \le 1$. The monotonicity property clearly holds. □

The primal protocol is indeed a MIMD (multiplicative increase, multiplicative decrease) protocol and like many protocols of this type converges fast.

Lemma 5. *For primal protocol $\mathcal{P}_\gamma$, the value of $t_{\max}(\mathcal{P}_\gamma)$ is of $O\left(\frac{1}{\gamma}\log\rho\right)$.*

Proof. Suppose all utilities except for the agent i are held fixed. The maximum time that the utility is less than $\frac{1}{2n}$ before reaching stable point is $\frac{1}{\gamma}$. Now suppose $x_i \ge \frac{1}{2n}$ at time zero. Note that the stable point of the utility is at most $c_{\max}$. Also, note that $\sigma(w_i)$ remains constant before reaching to the stable point.

There are two cases: if $\sigma(w_i) = 1$, by the Equation (4), $x_i(t) \ge \frac{1}{2n}e^{\gamma t}$. Similarly, for the case of $\sigma(w_i) = -1$, $x_i(t) \le c_{\max}e^{-\gamma t}$. Therefore, the total convergence time is at most $\frac{1}{\gamma}(\log 2nc_{\max} + 1)$. □

The corollary below is immediate form the lemma above and theorems 3.

Corollary 6. *The convergence time of the primal protocol $\mathcal{P}_\gamma$ into the equilibrium is $O\left(\frac{1}{\gamma}n\log\rho\right)$.*

4 A Dual Protocol

We now define a protocol in which the rate of change in the utility of an agent is (mostly) proportional to its dual price. We call this the *dual protocol*. Let $\xi > 0$ be a small constant. The rate of change in the utility of agent i, $1 \le i \le n$, is determined by the differential equation below:

$$\frac{dx_i}{dt} = -\frac{1}{\eta}\log w_i + \sigma(w_i)\xi. \tag{5}$$

When the utility is far from its stable point, $|\log w_i|$ is large and the utility converges to the stable points fast. Once it gets close of its stable point, $|\log w_i|$ is almost zero, ξ perturbs the rate of convergence and the utility smoothly reaches its stable point. The rate of changes in the utility around its stable point is very small, which makes this protocol appropriate for discrete implementation. We will now give an an upper bound on the convergence time.

Lemma 7. *For the dual protocol $\mathcal{P}$, the value $t_{\max}(\mathcal{P})$ is at most $c_{\max} + \frac{1}{\rho^2 \xi}$.*

Proof. Suppose all of the utilities except for the agent i are held fixed. Let x_i and x_i^* be the current and stable utility for agent i. First assume $x_i = x_i^* + \Delta$ and $\Delta > 0$. Also, assume the sum of the truncated amount of other utilities on resource e is y_j, i.e. $y_j = \sum_{k \neq i : j \in S_k} a_{kj} \min\{x_i, x_k\}$. Similarly, define $y_j^* = \sum_{k \neq i : j \in S_k} a_{kj} \min\{x_i^*, x_k\}$. For the dual price for agent i we have:

$$
\begin{aligned}
w_i &= \sum_{j \in S_i} a_{ij} \mu^{\eta \Lambda_j^i - 1} \\
&= \sum_{j \in S_i} a_{ij} \mu^{\eta \Delta / c_j + \eta (x_i^* + y_j)/c_j - 1} \\
&\geq \mu^{\eta \Delta / c_{\max}} \sum_{j \in S_i} a_{ij} \mu^{\eta (x_i^* + y_j)/c_j - 1} \\
&\geq \mu^{\eta \Delta / c_{\max}} \sum_{j \in S_i} a_{ij} \mu^{\eta \left(x_i^* + y_j^*\right)/c_j - 1} \\
&\geq \mu^{\eta \Delta / c_{\max}}
\end{aligned}
$$

The last equality follows from the fact that at the stable point the dual price is equal to 1; also, y_j^* is less than or equal to y_j. By the inequality above, $\frac{d\Delta}{dt} < -\frac{1}{\eta} \log w_i \leq \frac{-\log \mu}{c_{\max}} \Delta$. Similarly, for $\Delta < 0$, $-\frac{1}{\eta} \log w_i \geq \frac{-\log \mu}{c_{\max}} \Delta$ which yields $\frac{\dot{\Delta}}{\Delta} \leq \frac{-\log \mu}{c_{\max}}$. Therefore, after t time, $|\Delta| < c_{\max} \mu^{\frac{-t}{c_{\max}}}$. Consequently, after $c_{\max}$ time $|\Delta| \leq \frac{1}{\rho^2}$. Recall that before convergence, $|dx_i/dt| \geq \xi$. Hence, after $c_{\max} + \frac{1}{\rho^2 \xi}$ time, i becomes stable. □

The corollary below is immediate form the lemma above and the Theorem 3.

Corollary 8. *Analogous to the primal protocol, the dual protocol converges to the equilibrium in at most $O(nc_{\max})$ time.*

4.1 A Faster Dual Protocol

Sometimes, there is a large variance between the capacities of different resources. This variance has a major effect on the performance of the dual protocol defined in the previous section. In this section we adapt the dual protocol for such settings. Define $c_{\min}^i$ as the the minimum capacity in S_i, i.e. $c_{\min}^i = \min_{j \in S_i} \{c_j\}$. We define a protocol in which the rate of change in utility of agent i, $1 \leq i \leq n$, is determined by the differential equation

$$\frac{dx_i}{dt} = -c_{\min}^i \log w_i + \sigma(w_i) \xi.$$

We call this the *fast dual protocol*. In addition, we assume the initial point of x_i is at most $c_{\min}^i$.

Theorem 9. *Let r be the maximum number of agents that share a single resource. The fast dual protocol converges into the equilibrium in $O\left(nr\left(1 + \frac{1}{\xi \log \rho}\right)\right)$ time.*

Proof. We will now show that $t_{\max} = O\left(r\left(1 + \frac{1}{\xi \log \rho}\right)\right)$ for the fast dual protocol; the theorem will then follow from Theorem 3.

First, we show that $\mu^{\eta \Lambda_j^i - 1} \geq \frac{1}{\rho^2}$ implies that $c_j \leq 3\eta r c_{\min}^i$:

$$
\begin{aligned}
& \mu^{\frac{\eta r c^i_{\min}}{c_j}-1} \geq \mu^{\eta \Lambda^i_j - 1} \geq \tfrac{1}{\rho^2} \\
\Rightarrow \quad & \tfrac{\eta r c^i_{\min}}{c_j} - 1 \geq \log_\mu \tfrac{1}{\rho^2} = \tfrac{-2}{3} \\
\Rightarrow \quad & 3\eta r c^i_{\min} \geq c_j
\end{aligned}
$$

We use the definitions in Lemma 7. Let $\kappa = 3\eta r c^i_{\min}$. For $\Delta > 0$ we have:

$$
\begin{aligned}
w_i &= \sum_{j \in S_i} a_{ij} \mu^{\eta \Lambda_j - 1} \\
&= \sum_{j \in S_i} a_{ij} \mu^{\eta \Delta / c_j + \eta (x^*_i + y_j)/c_j - 1} \\
&\geq \sum_{j \in S_i, c_j \leq \kappa} a_{ij} \mu^{\eta \Delta / c_j + \eta (x^*_i + y_j)/c_j - 1} \\
&\geq \mu^{\eta \Delta / \kappa} \sum_{j \in S_i, c_j \leq \kappa} a_{ij} \mu^{\eta (x^*_i + y_j)/c_j - 1} \\
&\geq \mu^{\eta \Delta / \kappa} \left(1 - \frac{1}{\rho}\right)
\end{aligned}
$$

The last inequality is derived from the fact that the total contribution of the resources with capacity more than κ is less than $\frac{m}{\rho^2}$. By the inequality above:

$$
\begin{aligned}
\frac{d\Delta}{dt} &< -c^i_{\min} \log w_i \\
&\leq -c^i_{\min} \left(\frac{\eta \log \mu}{\kappa} \Delta + \log \left(1 - \frac{1}{\rho}\right) \right) \\
&= \frac{-\log \mu}{3r} \Delta - c^i_{\min} \log \left(1 - \frac{1}{\rho}\right)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad \Delta &< c^i_{\min} \mu^{\frac{-t}{3r}} - 3 r c^i_{\min} \frac{\log\left(1 - \frac{1}{\rho}\right)}{\log_\mu} \\
&= c^i_{\min} \mu^{\frac{-t}{3r}} + O\left(\tfrac{r}{\log \rho}\right)
\end{aligned} \tag{6}
$$

The last inequality follows by the Taylor expansion of $\log\left(1 - \frac{1}{\rho}\right)$. Similar results holds for $\Delta < 0$. Therefore, after $O(r)$ time, $|\Delta|$ becomes $O\left(\frac{r}{\log \rho}\right)$. After that, the rate of convergence is at least ξ, so the utility of agent i settles into the equilibrium in $O\left(r + \frac{r}{\xi \log \rho}\right)$ time. □

5 Bandwidth allocation on a single link

Consider the case of a single resource and n agents, and $a_{i1} = 1$, for $1 \leq i \leq n$. This is equivalent to the bandwidth allocation on a single link and we can prove faster convergence results.

Proposition 10. *Any truncated protocol over a single link network converges in at most $2t_{\max}$ time.*

Proof. By symmetry, at equilibrium, all of the utilities are equal. Let f be the utility at the equilibrium. Similar to the proof of Theorem 3, there are two phases. In the first phase, all utilities become at least f. In the second phase, all utility decrease to become stable at f. □

Corollary 11. *The convergence of the primal protocol on a single link is logarithmic in n. Also, the dual protocol converges into the equilibrium in $O(1)$ time.*

Open Problems

We conclude by describing some open problems:

1. Can our results be extended to the case where individual utilities are arbitrary concave functions of the allocated resources?
2. Can our techniques be applied to exchange economies, where the resources are also held by agents who use the revenue acquired from allocating their resources to acquire a share of other resources?
3. Our price based protocols require each resource to know individual allocations made to each agent. Can an aggregate measure be used instead?

References

[1] Yossi Azar, Leah Epstein, Yossi Richter, and Gerhard J. Woeginger. All-norm approximation algorithms. *J. Algorithms*, 52(2):120–133, 2004.

[2] Yossi Azar and Shai Taub. All-norm approximation for scheduling on identical machines. In *Scandinavian Workshop on Algorithm Theory*, pages 298–310, 2004.

[3] Y. Bartal, J. Byers, and D. Raz. Global optimization using local information with applications to flow control. *38th Annual Symposium on Foundations of Computer Science*, pages 303–312, 1997.

[4] R. Bhargava, A. Goel, and A. Meyerson. Using approximate majorization to characterize protocol fairness. *Proceedings of ACM Sigmetrics*, pages 330–331, June 2001.

[5] N. Buchbinder and J. Naor. Fair online load balancing. *Proceedings of the 18th Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 291–298, 2006.

[6] N. Buchbinder and J. Naor. Improved bounds for online routing and packing via a primal-dual approach. *Proceeding of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 293–304, 2006.

[7] Ashok K. Chandra and C. K. Wong. Worst-case analysis of a placement algorithm related to storage allocation. *SIAM J. Comput.*, 4(3):249–263, 1975.

[8] X. Chen and X. Deng. Settling the complexity of two-player nash equilibrium. *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

[9] S. Cho and A. Goel. Bandwidth allocation in networks: a single dual-update subroutine for multiple objectives. *Proceedings of the first Workshop on Combinatorial and Algorithmic Aspects of Networks*, 3405:28–41, 2004.

[10] S. Cho and A. Goel. Pricing for fairness: distributed resource allocation for multiple objectives. *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 197–204, May 2006.

[11] B. Codenotti, B. McCune, S. Penumatcha, and k. Varadarajan. Market equilibrium for ces exchange economies: Existence, multiplicity, and computation. *FSTTCS*, 2005.

[12] B. Codenotti, A. Saberi, K. Varadarajan, and Y. Ye. Leontief economies encode nonzero sum two-player games. *In the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[13] C. Daskalakis, P. Goldberg, and C. Papadimitriou. The complexity of computing a nash equilibrium. *Proceedings of the 38th ACM Annual Symposium on Theory of Computing*, 2006.

[14] N. Devanur, C. Papadimitriou, A. Saberi, and V. Vazirani. Market equilibrium via a primal-dual-type algorithm. *Proceedings of the 44th IEEE Symposium on the Foundations of Computer Science*, pages 389–395, 2002.

[15] N. Garg and J. Konemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. *39th Annual Symposium on Foundations of Computer Science*, pages 300–309, 1998.

[16] A. Goel and A. Meyerson. Simultaneous optimization via approximate majorization for concave profits or convex costs. *Algorithmica*, 44(4):301–323, May 2006.

[17] A. Goel, A. Meyerson, and S. Plotkin. Combining fairness with throughput: Online routing with multiple objectives. *Journal of Computer and Systems Sciences*, 63(1):62–79, 2001.

[18] A. Goel, A. Meyerson, and S. Plotkin. Approximate majorization and fair online load balancing. *ACM Transactions on Algorithms*, 1(2):338–349, Oct 2005.

[19] G.H. Hardy, J.E. Littlewood, and G. Polya. Some simple inequalities satisfied by convex functions. *Messenger Math.*, 58:145–152, 1929.

[20] G.H. Hardy, J.E. Littlewood, and G. Polya. *Inequalities.* 1st ed., 2nd ed. Cambridge University Press, London and New York., 1934, 1952.

[21] F.P. Kelly, A.K. Maulloo, and D.K.H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[22] J. Kleinberg, Y. Rabani, and E. Tardos. Fairness in routing and load balancing. *J. Comput. Syst. Sci.*, 63(1):2–20, 2001.

[23] A. Kumar and J. Kleinberg. Fairness measures for resource allocation. *Proceedings of 41st IEEE Symposium on Foundations of Computer Science*, 2000.

[24] S. Low, L. Peterson, and L. Wang. Understanding TCP Vegas: a duality model. *Proceedings of ACM Sigmetrics*, 2001.

[25] A.W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications.* Academic Press (Volume 143 of Mathematics in Science and Engineering), 1979.

[26] S. Plotkin, D. Shmoys, and E. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Math of Oper. Research*, pages 257–301, 1994.

[27] Y. Ye. A note on exchange market equilibria with leontief's utility: Freedom of pricing leads to rationality. *Proceeding of The 1st International Workshop On Internet and Network Economics*, 2005.

A Majorization

It is known ([16, 19]) that the problem of approximating uncountably many canonical utility functions can be reduced to approximating n "prefix" functions. Define the k-th prefix, $p_k(\mathbf{x})$ to be the sum of the k smallest components of $\mathbf{x}$ (not $x_1 + x_2 + \ldots + x_k$ but $x_{\sigma(1)} + x_{\sigma(2)} + \ldots x_{\sigma(k)}$ where σ is the permutation that sorts $\mathbf{x}$ in increasing order). Let P_k^* denote the maximum possible value of $p_k(\mathbf{x})$ subject to capacity constraints. A feasible solution $\mathbf{x}$ is said to be α-majorized if $\alpha \geq \frac{P_k^*}{p_k(\mathbf{x})}$ for all $1 \leq k \leq n$.

Theorem 12. *[19, 16] A feasible solution* $\mathbf{x}$ *is* α*-fair if and only if* $\mathbf{x}$ *is* α*-majorized.*

B Proof of Theorem 2

Let $\mathbf{x}$ be the allocation found by distributed-weighted-majorized algorithm. For $1 \leq i \leq n$, let P_i be the i-th prefix sum $\mathbf{x}$. We show $P_i \log \mu \geq P_i^*$, where P_i^* is the maximum weighted prefix sum subject to feasibility. This implies that $\mathbf{x}$ is $\log \mu$-majorized. Therefore, by Theorem 12, $\mathbf{x}$ is $\log \mu$-fair.

The main idea of the proof is based on proposition below and is immediate from [10]. In the sequel, we use l to denote the vector of dual prices and D_l as $\sum_j l_j c_j$.

Proposition 13. *For every dual price l, let $\beta_k(l) = \min_{\mathbf{x}} \frac{\sum_i \widetilde{w}_i x_i}{P_k(\mathbf{x})}$ where $\widetilde{w} = \min\{1, w\}$. Then, for any feasible allocation $\mathbf{x}$, $P_k(\mathbf{x}) \leq \frac{D_l}{\beta_k(l)}$.*

Let T be the first time $\beta_i(l) = 1$. We show for $0 \leq t \leq T$, $D_l(t) \leq \log \mu P_i(t)$:

$$\begin{aligned}
\frac{dD_l}{dt} &= \frac{d}{dt} \sum_j \mu^{\eta \Lambda_j - 1} c_j \\
&= \sum_j \eta \log \mu \frac{d\Lambda_j}{dt} \mu^{\eta \Lambda_j - 1} c_j \\
&= \eta \log \mu \sum_j \left(\sum_{i: w_i < 1} a_{ij} \gamma \right) \mu^{\eta \Lambda_j - 1} \\
&= \eta \log \mu \sum_{i: w_i < 1} w_i \gamma
\end{aligned} \tag{7}$$

In [10], Cho and Goel proved the lemma below:

Lemma 14. *Given a weight vector w, let k denote the cardinality of the set $\{w_j : w_j \geq 1\}$. If $\beta_i(l) \leq 1$ for some i, then $\sum_{j: w_j < 1} w_j \ \leq \ i - k$.*

This lemma still holds for the Leontief utilities. By this lemma and inequality (7):

$$\frac{dD_l}{dt} \leq \eta \log \mu \sum_{i: w_i < 1} w_i \gamma \leq \eta \log \mu \frac{d}{dt} P_i(x)$$

Taking the integral from both sides:

$$\begin{aligned}
D_l(T) &\leq \eta \log \mu P_i(T) - \eta \log \mu P_i(0) + \eta D_l(0) \\
&\leq \eta \log \mu P_i(T) - \eta \frac{i \log \mu}{\rho} + \frac{m c_{\max}}{\rho^3} \\
&\leq \eta \log \mu P_i(T)
\end{aligned}$$

At time T, $\beta_i(l) = 1$; $P_k^* \leq D_l(T) \leq \eta \log \mu P_i(T)$. Also, P_i is increasing over time. Therefore, at the end of the algorithm, $P_i^* \leq \log \mu P_i$, which completes the proof.