Computer Networks 52 (2008) 2947–2960

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Congestion control in utility fair networks

Tobias Harks *, Tobias Poschwatta

Technische Universität Berlin, COGA, Straße des 17. Juni 136, 10623 Berlin, Germany

ARTICLE INFO

Article history:
Received 10 August 2007
Received in revised form 27 May 2008
Accepted 19 June 2008
Available online 28 June 2008

Responsible Editor: N. Akar

Keywords:
Utility fairness
Congestion control
Non-convexity
Multimedia

ABSTRACT

This paper deals with a congestion control framework for elastic and real-time traffic, where the user's application is associated with a utility function. We allow users to have concave as well as non-concave utility functions, and aim at allocating bandwidth such that utility values are shared fairly. To achieve this, we transform all utilities into strictly concave second order utilities and interpret the resource allocation problem as the global optimization problem of maximizing aggregate second order utility. We propose a new fairness criterion, *utility proportional fairness*, which is characterized by the unique solution to this problem. Our fairness criterion incorporates utility max–min fairness as a limiting case. Based on our analysis, we obtain congestion control laws at links and sources that (i) are linearly stable regardless of the network topology, provided that a bound on round-trip-times is known, (ii) provide a utility proportional fair resource allocation in equilibrium. We further investigate the *efficiency* of utility fair resource allocations. Our measure of efficiency is defined as the worst case ratio of the total utility of a utility proportional fair rate vector and the maximum possible total utility. We present a generic technique, which allows to obtain upper bounds on the efficiency loss. For special cases, such as linear and concave utility functions, and non-concave utility functions with bounded domain, we explicitly calculate such upper bounds. Then, we study utility fair resource allocations with respect to bandwidth fairness. We derive a fairness metric assessing the aggressiveness of utility functions. This allows us to design fair utility functions for various applications. Finally, we simulate the proposed algorithms using the NS2 simulator.

1. Introduction

In the last years, congestion control of communication networks has been interpreted as a distributed algorithm at sources and links in order to solve a global optimization problem [1–4]. Each user is associated with an increasing, strictly concave bandwidth utility function representing elastic traffic. The congestion control algorithms aim at maximizing aggregate utility subject to capacity constraints on the links. The solution to this problem is derived by decomposing the overall problem into subproblems that can be solved by links and sources using only local information. The links communicate a price based on usage measurements; the source collects the aggregate price on its path and adapts its sending rate in order to maximize its surplus.

Even though considerable progress has been made in this direction, the existing work focuses only on elastic traffic, such as file transfer (FTP, HTTP) or electronic mail (SMTP). As shown in [5], some applications, especially real-time applications, have non-concave bandwidth utility functions. It is known, however, that the afore mentioned distributed algorithms do not maximize non-concave aggregate utility in general, see for instance Lee et al. [6]. Furthermore, it is easy to construct non-concave instances, where a system optimal solution distributes bandwidth such that only a few applications share the available bandwidth, while the rest receives zero utility (bandwidth).

* Corresponding author. Tel.: +49 30 206187917.
E-mail address: harks@math.tu-berlin.de (T. Harks).

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.06.013

In this regard, several works (among others Sarkar and Tassiulas [7], Cao and Zegura [8], and Liao and Campbell [9]) argue that it is an application performance measure, i.e., the utility that should be shared fairly among users. A user running an application does not care about any fair bandwidth shares as long as his application performs satisfactory. To illustrate this paradigm, let us consider a single link of capacity c shared by two users. One user transfers data according to an elastic application with strictly increasing, and concave bandwidth utility $U_1(\cdot)$. The other user transfers real-time video data with a non-concave bandwidth utility function $U_2(\cdot)$ (steps represent encoding layers).

Fig. 1 shows, how different bandwidth allocations affect the received utility. If bandwidth is shared equally, which corresponds to a max–min fair bandwidth allocation, user 1 receives a much larger utility than user 2. In fact, user 2 would receive a utility value of zero, that is, the minimum encoding rate is not achieved. If we want to share *utility* equally instead of bandwidth, we would like to have a resource allocation, where the received utilities are equal or *utility max–min fair*, i.e., $U_1(x_1) = U_2(x_2) = u^*$.

There are several challenges in designing utility fair congestion control algorithms that have not been fully solved so far. Is it possible to design distributed and stable algorithms converging to some utility fair operating point under relaxed assumptions on utility functions? Can we characterize the efficiency of utility fair rate allocations in terms of system utility? Can we design utility functions such that the resulting rate allocations are bandwidth fair in the long term? After reviewing the related work, we outline our contributions, which partially answer these questions.

1.1. Related work

The existing work on congestion control algorithms using the utility framework is focused on elastic traffic such as TCP. Congestion control mechanisms are regarded as a distributed algorithm carried out by sources and links in order to solve a global optimization problem, see [1–4] or the book of Srikant [10] and references therein. The objective is to maximize aggregate source utility over transmission rates subject to capacity constraints:

$$\max_{x_s \in X_s} \sum_{s \in S} U_s(x_s) \text{ s.t. } \sum_{s \in S(l)} x_s \leqslant c_l, l \in L. \quad (1)$$

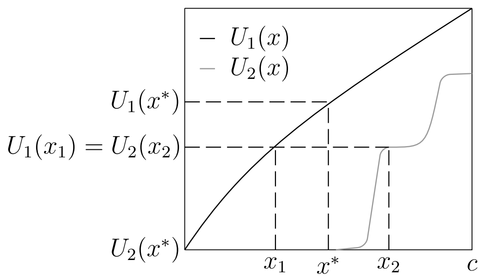


Fig. 1. Utility functions with $U_1(x) < U_2(x), x \in [0, c]$. The rate allocation $y_1 = y_2 = x^*$ is max–min fair, whereas the rate allocation x_1 and x_2 are utility max–min fair.

Source rates x_s can be interpreted as primal variables, congestion measures p_l as dual variables. Using the dual approach, a gradient projection method to generate optimal prices is applied to the dual objective function, see [1]:

$$\dot{p}_l = \begin{cases} \frac{1}{c_l}(y_l - c_l), & \text{if } p_l > 0, \\ \frac{1}{c_l}[y_l - c_l]^+, & \text{if } p_l = 0, \end{cases} \quad (2)$$

where y_l denotes the aggregate sending rate on link l. In [11], it is shown that if the utility functions are strictly concave (2), combined with the dynamic source law

$$\tau_s \dot{\xi}_s = \beta_s (U_s'(x_s) - q_s), x_s = x_s^{\max} \exp\left(\xi_s - \frac{\alpha_s q_s}{M_s \tau_s}\right). \quad (3)$$

converges to the unique optimal solution $x_s^* = U_s'^{-1}(q_s^*)$ starting from any initial condition (M_s is an upper bound on the number of bottlenecks, α_s and β_s are positive parameters). Furthermore, this approach has the appealing property that the equilibrium is locally stable within given delay bounds [12]. Following [4], we can associate a class of concave utility functions with corresponding bandwidth fairness-criteria as follows:

$$U_s(x_s, \eta_s) = \begin{cases} -w_s \frac{x_s^{1-\eta_s}}{1-\eta_s}, \eta_s > 0, \ \eta_s \neq 1, \\ w_s \log(x_s), \eta_s = 1. \end{cases} \quad (4)$$

Here, the values w_s are fixed arbitrary nonnegative values. Then, in the case $\eta_s = 1$, we have *weighted proportional* fairness [2]. In the case $\eta_s = 2$, we have *minimum potential delay* fairness, and for $\eta_s \to \infty$, we have *max–min* fairness.

Another line of research focuses on utility functions that are not necessarily strictly concave. As illustrated in Fig. 1 utilities describing real-time traffic are in general non-concave. We can classify the work in this area into two categories:

(i) approaches that aim at maximizing aggregate (non-concave) utility subject to capacity constraints: due to the possible duality gap of this type of problem, stable decentralized algorithms can be derived only in special cases. Lee, Mazumdar and Shroff showed in [6] that the canonical distributed algorithms (3) and (2) may fail to converge to a feasible rate allocation and may lead to instability and congestion. To overcome these problems they proposed a 'self regulating' heuristic for the special type of sigmoidal utilities in combination with a sub-gradient method to generate prices. Chiang, Zhang and Hande [13] examined the conditions under which the canonical algorithms still converge to the globally optimal rate despite the non-concavity of utility functions. Fazel and Chiang [14] presented heuristics to maximize non-concave utility based on the sum of squares method. Their heuristics, however, provide no quality guarantee in the worst case.

(ii) approaches that are concerned with a different fairness definition rather than maximizing aggregate utility: an equilibrium point should result in roughly equal utility values for different applications [7–9,15]. In [7], only mild assumptions on the feasible utility functions are required (non-decreasing, not necessarily continuous, min. bandwidth exists for a given utility value). The drawbacks of this approach are that the links have to maintain per-flow states in order to allocate bandwidth utility fair, and that there are no stability results given in the presence of communication delay. Cao and Zegura presented in [8] a link algorithm that achieves a utility max–min fair bandwidth allocation, where for each link the utility functions of all flows sharing that link is maintained. Hyang and Song [16] and Cho and Song [17] presented distributed algorithms without per-flow states that converge to a utility max–min fair operating point. However, they proved stability under bounded communication delay only in the single link case. More recently, based on the work of Zhang et al. [18], Miller and Harks [19] presented distributed utility max–min fair algorithms that are locally stable in the presence of time-varying delays.

1.2. Our contribution

First, we align the convex optimization framework with the utility fairness approach. We achieve this by transforming all (possibly non-concave) utility functions into strictly concave *second order utility functions*. Then, analogous to (1), we interpret an operating point as the solution to maximizing second order utility subject to capacity constraints. This approach enables us to use scalable, decentralized, and stable congestion control algorithms in the line of [1,3,12]. Yet, we relax the concavity assumption on the bandwidth utilities, and achieve utility fairness in equilibrium. We further define a new fairness criterion, *utility proportional fairness* that includes utility max–min as a limiting case. We emphasize that the distributed algorithms do not need any per-flow information at the links, and are locally stable regardless of the network topology provided a bound on the round-trip-time is known. The feedback from links to sources only relies on the communication of Lagrange multipliers, called shadow prices, from the links to the sources. This can be achieved by an active queue management (AQM) scheme using explicit congestion notification (ECN) [20].

Second, we investigate the *efficiency* of utility fair resource allocations. Our measure of efficiency (termed *price of fairness*) is defined as the worst case ratio of the total utility of a utility proportional fair rate vector and the maximum possible total utility. We present a generic technique to obtain upper bounds on the price of fairness. For special cases, such as linear and concave utility functions, and non-concave utility functions with bounded domain, we explicitly calculate upper and lower bounds.

Third, we study utility fair resource allocations with respect to bandwidth fairness. In general, the resulting bandwidth allocation of a utility fair operating point strongly depends on the bandwidth utility functions that are used. In the existing literature on utility fair networks, e.g. [7–9,16,21,22], a fairness metric for utility functions is missing. We address this issue by defining a fairness measure for bandwidth utility functions. Based on the induced fairness metric, we are able to design normalized utility functions with equal expected bandwidth shares.

Most of the results of this paper have previously been published by Harks [15] and Harks and Poschwatta [23].

1.3. Paper organization

The rest of the paper is organized as follows. In Section 2, we introduce the basic model and our congestion control framework. Then, in Section 3, we define our new fairness criterion, which we term *utility proportional fairness*. We show that our algorithms achieve utility max–min fairness for the entire network as a limiting case. In Section 4, we study the efficiency of utility proportional fair resource allocations with respect to maximum possible total utility. We derive in Section 5 a fairness measure for bandwidth utilities that allows to assess the aggressiveness of different utilities and serves as a design framework for fair utility functions. Finally, we evaluate the derived congestion control algorithms via simulations in Section 6.

2. Transformation and convex optimization

We model a capacitated packet switched network $N = (V, L, c)$ by a set of nodes (routers) V connected by a set L of unidirectional links (output ports) with finite capacities $\vec{c} = (c_l, l \in L)$. The set of links is shared by a set S of sources indexed by s. A source s represents an end-to-end connection and its route involves a subset $L(s) \subseteq L$ of links. Equivalently, each link l is used by a subset $S(l) \subseteq S$ of sources. For notational convenience we sometimes use R to denote the routing matrix for a given directed network with a set of sources. The tuple $I = (N, R)$ completely characterizes the underlying network and sources with corresponding routes. We use $\mathscr{I}$ to denote a class of tuples I, where each $I \in \mathscr{I}$ corresponds to a network with a routing matrix for a given set of sources. As an example of such a class consider the class $\mathscr{I}'$ of single link networks.

A transmission rate $x_s(t) \in X_s = [0, x_s^{\max}]$ in *packets per second* is associated with each source s. A rate vector $\vec{x}(t) = (x_s(t), s \in S)$ is said to be *feasible* if it satisfies the conditions: $x_s(t) \in X_s \forall s \in S$ and $\sum_{s \in S(l)} x_s(t) \leqslant c_l \forall l \in L$. Whenever we refer to a steady state rate vector, we only write $\vec{x}$ and omit the time dependency. With each link l, a scalar positive congestion measure $p_l(t)$, called *price*, is associated. Let $y_l(t) = \sum_{s \in S(l)} x_s(t - \tau_{ls}^f)$ be the aggregate transmission rate of link l, i.e. the sum of all rates using that link in which the forward delays τ_{ls}^f between sources and links are accounted for. Let $q_s(t) = \sum_{l \in L(s)} p_l(t - \tau_{ls}^b)$ be the end-to-end congestion measure of source s, where again τ_{ls}^b are the backward delays from links to sources. The total RTT is given by $\tau_s = \tau_{ls}^f + \tau_{ls}^b$. If the transmission rate of source s is x_s, source s receives a benefit measured by a continuous, nondecreasing bandwidth utility function $U_s(x_s)$. We denote by $\vec{U}$ the vector of utility functions $U_s, s \in S$. In the following, we describe a constructive method to transform possibly non-concave utility functions into appropriate strictly concave functions, which we term *second order utilities*. We will define an equilibrium point as the unique maximizer of aggregated second order utility. Subsequently, we will show that the resulting resource allocation provides a fair share of an application layer performance measure, i.e. the bandwidth utility to users. In contrast to [1–3,11,24–27], we do not pose any restrictions on the bandwidth utility functions, except for monotonicity.

2.1. Bandwidth utilities and transformation function

If the transmission rate of user s is x_s, user s receives a benefit measured by the bandwidth utility $U_s(x_s)$.

Assumption 1. A bandwidth utility function $U_s(x_s)$ is feasible, if it satisfies:

(1) $U_s(x_s) \geqslant 0, U_s(0) =: u_s^{\min} \geqslant u_{\min}, U_s(x_s^{\max}) =: u_s^{\max} \leqslant u_{\max}$,
(2) $U_s'(x_s) \geqslant \gamma_1 > 0, U_s'(x_s) \leqslant \gamma_2 < \infty$, for nonnegative constants γ_1 and γ_2, where $U'(x)$ is defined as $\frac{dU(x)}{dx}$.

We denote a class of feasible bandwidth utility functions by $\mathscr{U}$.

Note that we do not rely on concavity.

The first monotonicity assumption $U_s'(x_s) \geqslant \gamma_1$ ensures the existence of the inverse function $U_s^{-1}(\cdot)$ over the range $Y_s := [u_s^{\min}, u_s^{\max}]$. The second assumption $U_s'(x_s) \leqslant \gamma_2 < \infty$ is required to ensure stability of the congestion control algorithms that are presented in Section 2.3. Before we present a constructive method to generate second order utility functions, we briefly restate the overall paradigm. An optimal operation point or equilibrium should result in almost equal utility values for different applications. The exact definition of the proposed resource allocation, i.e. *utility proportional fair* resource allocation, will be given below. To follow this paradigm, we translate a given congestion level of a path, represented by q_s, into an appropriate utility value that the network can offer to source s. We model this utility value, which we term *available utility*, as the transformation of the congestion measure q_s by a *transformation function* $f_s(q_s)$. This function is assumed to be strictly decreasing.

Assumption 2. The transformation function $f_s(\cdot)$ describing the available utility of a path used by sender s is assumed to be a continuously differentiable and strictly decreasing function of the aggregate congestion measure q_s, i.e., $f_s'(q_s) \leqslant -\gamma_3, \gamma_3 > 0$, for all $q_s \geqslant 0$ and $s \in S$. We denote by $\vec{f}$ the vector of transformation functions $f_s(\cdot), s \in S$.

The monotonicity assumption is reasonable, since the more congested a path is, the smaller will be the available utility for an application. The main idea is that each user s should send at data rates x_s in order to match its own bandwidth utility with the available utility of its path. This leads to the following equation:

$$U_s(x_s) = [f_s(q_s)]_{u_s^{\min}}^{u_s^{\max}}, \quad s \in S, \tag{5}$$

where

$$[w]_a^b := \min\{\max\{w, a\}, b\} = \begin{cases} w, & \text{if } a \leqslant w \leqslant b, \\ a, & \text{if } w < a, \\ b, & \text{if } w > b. \end{cases}$$

Note that the utility functions for source s is bounded by the minimum and maximum utility values $u_s^{\min}$ and $u_s^{\max}$. Hence, the source rates x_s are adjusted according to the available utility $f_s(q_s)$ of their used path as follows:

$$x_s = U_s^{-1}\left([f_s(q_s)]_{u_s^{\min}}^{u_s^{\max}}\right), \quad s \in S. \tag{6}$$

A source $s \in S$ reacts to the congestion measure q_s in the following manner: if the congestion measure q_s is below a threshold $q_s < q_s^{\min} := f_s^{-1}(u_s^{\max})$, then the source transmits data at maximum rate $x_s^{\max} = U_s^{-1}(u_s^{\max})$; if q_s is above a threshold $q_s > q_s^{\max} := f_s^{-1}(u_s^{\min})$, the source sends at minimum rate $0 = U_s^{-1}(u_s^{\min})$; if q_s is in between these two thresholds $q_s \in Q_s := [q_s^{\min}, q_s^{\max}]$, the sending rate is adapted according to $x_s = U_s^{-1}(f_s(q_s))$.

Lemma 1. *The function $G_s(q_s) = U_s^{-1}([f_s(q_s)]_{u_s^{\min}}^{u_s^{\max}})$ is nonnegative, differentiable, and strictly decreasing, i.e., $G_s'(q_s) < 0$, on the interval $q_s \in Q_s$, and its inverse $G_s^{-1}(\cdot)$ is well defined on X_s.*

Proof. Since $U_s(\cdot)$ is nonnegative on X_s, the inverse $U_s^{-1}(\cdot)$ is also nonnegative on Y_s. Using that $f_s(\cdot)$ is differentiable over Q_s, and $U_s^{-1}(\cdot)$ is differentiable over Y_s, the composition $G_s(q_s) = U_s^{-1}(f_s(q_s))$ is differentiable over Q_s. We compute the derivative using the chain rule: $G_s'(q_s) = U_s^{-1'}(f_s(q_s))f_s'(q_s)$. The derivative of the inverse $U_s^{-1}(f_s(q_s))$ can be computed as

$$U_s^{-1'}(f_s(q_s)) = \frac{1}{U_s'(U_s^{-1}(f_s(q_s)))} > 0.$$

With the inequality $f_s'(\cdot) < 0$, we get $G_s'(q_s) < 0$, for all $q_s \in Q_s$. Hence, $G_s(q_s)$ is strictly monotone decreasing in Q_s, so its inverse $G_s^{-1}(x_s)$ exists on X_s. □

2.2. Equilibrium structure and second order utility optimization

We study the above model at equilibrium, i.e., we assume that rates and prices are at fixed equilibrium values $\vec{x}^*, \vec{y}^*, \vec{p}^*, \vec{q}^*$. In equilibrium, the sending rates satisfy:

$$x_s^* = U_s^{-1}\left([f_s(q_s^*)]_{u_s^{\min}}^{u_s^{\max}}\right) = G_s(q_s^*), s \in S. \tag{7}$$

Since q_s represents the congestion in the path $L(s)$, the sending rates will be decreasing at higher q_s, and increasing at lower q_s. Now we consider the inverse $G_s^{-1}(x_s)$ of the above function on the interval X_s. Then, we construct the second order utility $F_s(x_s)$ as the integral of $G_s^{-1}(x_s)$. Hence, $F_s(\cdot)$ has the following form and property:

$$F_s(x_s) := \int G_s^{-1}(x_s)dx_s \text{ with } F_s'(x_s) = G_s^{-1}(x_s). \tag{8}$$

Lemma 2. *The second order utility $F_s(x_s)$ is a nonnegative, continuous, strictly increasing, and strictly concave function on the interval X_s.*

Proof. This follows directly from Lemma 1 and the relation

$$F_s''(x_s) = G_s^{-1'}(x_s) = \frac{1}{G_s'(q_s)} < 0. \qquad \square$$

The construction of $F_s(x_s)$ leads to the following property:

Lemma 3. *The equilibrium rate (7) is the unique solution of the optimization problem:*

$$\max_{x_s \in X_s} F_s(x_s) - q_s^* x_s. \tag{9}$$

Proof. The first order necessary optimality conditions for a solution x_s^* of problem (9) are:

$$x_s^* = \begin{cases} 0, & \text{if } F_s'(0) \leqslant q_s^*, \\ x_s^{\max}, & \text{if } F_s'(x_s^{\max}) \geqslant q_s^*, \\ U_s^{-1}(f_s(q_s^*)), & \text{else.} \end{cases}$$

Then, it follows that the source law (7)

$$x_s^* = U_s^{-1}\left([f_s(q_s^*)]_{u_s^{min}}^{u_s^{\max}}\right)$$

satisfies the above optimality conditions. Due to the strict concavity of $F(x_s)$ on X_s, the second order sufficient condition is also satisfied, completing the proof. □

2.3. Dual problem and stability

We interpret an equilibrium point as the unique solution of the following convex optimization problem:

$$\max_{x_s \in X_s, s \in S} \sum_{s \in S} F_s(x_s) \quad \text{s.t.} \quad R\vec{x} \leqslant \vec{c}. \tag{10}$$

As we will show in the next section, the optimal solution to (10) will ensure a general notion of utility fairness, which includes utility max–min fairness in a limiting case. Problem (10) can be regarded as an artificial optimization problem leading to scalable distributed algorithms that are utility fair in equilibrium. To solve (10) and its dual, we use the dynamic dual link law (2).

To ensure stability in presence of communication delays we insert the second order utility in the dynamic source law (3):

$$\tau_s \dot{\xi}_s = \beta_s (f_s^{-1}(U_s(x_s)) - q_s), x_s = x_s^{\max} \exp\left(\xi_s - \frac{\alpha_s q_s}{M_s \tau_s}\right). \tag{11}$$

This ensures that the elasticity (gain) of the static demand curve (6) does not affect the stability. At low gains $G_s'(q_s) \leqslant v$, for some small v, the static source law given in (6) would be sufficient to achieve stability; at high gains $G_s'(q_s) > v$, the dynamic in $\dot{\xi}$ forces the aggressive source to slow down in order to maintain stability.

We conclude this section by recalling a stability result of the dual algorithm (2) combined with the dynamic source law (11) due to [12].

Theorem 1. *Assume the routing matrix R is nonsingular and the round-trip-times are bounded, i.e. for every s: $\tau_s \leqslant \bar{\tau}$. Consider the source law (11). At equilibrium, the system will satisfy the desired demand curve (6). Furthermore, for β_s small enough and $\alpha_s < \frac{\pi}{2}$ the dual algorithm (2) in combination with (11) is locally stable.*

Note that for proving this result with techniques presented in [12], it is required that the second derivative of $F_s(x_s)$ is bounded away from zero. This is actually ensured by the conditions $U_s'(x_s) \leqslant \gamma_2$ and $f_s'(q_s) \leqslant -\gamma_3$, in Assumptions 1 and 2, respectively.

Remark 1. Our transformation of non-concave utility functions into strictly concave second order utility functions leads to a concave utility maximization problem, see (10). Thus, in order to achieve utility fairness in equilibrium, we can use any decentralized congestion control framework that maximize concave network utility, see for instance the many variants suggested in Srikant [10].

3. Utility proportional fairness

After we have established local stability of the distributed algorithms (2) and (11), we will characterize in this section the fairness conditions of the equilibrium. Before we come to our new fairness definition, we restate the concept of utility max–min fairness. It is simply the translation of the well-known bandwidth max–min fairness applied to utility values.

Definition 1. A rate vector $\vec{x}$ is said to be *utility max–min fair*, if it is feasible, and for any other feasible rate vector $\vec{y}$, the following condition hold: if $U_s(y_s) > U_s(x_s)$ for some $s \in S$, then there exists $k \in S$ such that $U_k(y_k) < U_k(x_k)$ and $U_k(x_k) \leqslant U_s(x_s)$.

Suppose we have a utility max–min fair rate allocation. Then, a user cannot increase its utility, without decreasing the utility of another user, who already receives a smaller utility. We now apply the above definition to a utility allocation of a single path.

Definition 2. Consider a single path in the network denoted by a set of links $(l \in L_p)$. Assume a set of users $S_{L_p} \subseteq S$ share this path, i.e. $L(s) = L_p$ for $s \in S_{L_p}$. Then, the rate vector $\vec{x}$ is said to be *path utility max–min fair* if the rate allocation $x_s, s \in S_{L_p}$ on every path L_p is utility max–min fair.

Note that the definition of path utility max–min fairness does not imply that every user using the same path L_p receives the same utility. Instead, it is possible that some users are already at their maximum rate receiving lower utility than others.

Now we come to our proposed new fairness criterion, based on the second order utility optimization framework.

Definition 3. Assume, all second order utilities $F_s(\cdot)$ are of the form (8). A rate vector $\vec{x}$ is called *utility proportional fair* if for any other feasible rate vector $\vec{y}$ the following optimality condition is satisfied:

$$\sum_{s \in S} F_s'(x_s)(y_s - x_s) = \sum_{s \in S} G_s^{-1}(x_s)(y_s - x_s) = \sum_{s \in S} f_s^{-1}(U_s(x_s))(y_s - x_s) \leqslant 0. \tag{12}$$

The above inequality is known as a type of *variational inequality*. It ensures, that any proportional utility fair rate vector will solve the optimization problem (10) and vice versa.

Lemma 4. *A rate vector $\vec{x}$ is utility proportional fair if and only if it solves problem (10).*

The proof simply uses the characterization of solutions of the convex problem (10) via the variational inequality (12). Using the above lemma, we can show that a utility proportional fair rate vector always exists and is unique.

Theorem 2. *A utility proportional fair rate vector always exists and is uniquely determined.*

Proof. Lemma 4 shows that every utility proportional fair rate vector solves problem (10). Since the objective function for (10) is continuous and the set of feasible rate vectors is compact, there exists an optimal solution. Furthermore, using Lemma (2), the objective function is strictly concave, thus, the optimal solution is unique. □

Remark 2. Problem (10) qualitatively falls into the class of well studied network utility maximization problems, see [1–4,24,25]. In particular, it follows that the solution vector $\vec{x}$ is Pareto[1] efficient, see Tang et al. [28]. Additionally, it can be shown that the vector of utilities $\vec{U}$ is also Pareto efficient. This is implied by the strict monotonicity of feasible utility functions.

If we assume that all users have the same transformation function, that is, $f(\cdot) = f_s(\cdot)$ for all $s \in S$, then we have the following properties of a utility proportional fair rate allocation, which are proven in Appendix A. Note that the case of a common transformation function for all users is of independent interest as it allows to communicate the feedback value $f(q)$ by routers. Thus, the only private information of end users is their utility function.

Theorem 3. *Suppose that all users have a common transformation function $f(\cdot)$ and all second order utility functions are defined by (8). Let the rate vector $\vec{x}$ be proportional utility fair, i.e. the unique solution of (10). Then the following properties hold:*

(i) The rate vector $\vec{x}$ is path utility max–min fair.
(ii) If $q_{s_1} \in Q_{s_1}, q_{s_2} \in Q_{s_2}$ and $q_{s_1} \leqslant q_{s_2}$ for sources s_1, s_2, then $U_{s_1}(x_{s_1}) \geqslant U_{s_2}(x_{s_2})$.
(iii) If source s_1 uses a subset of links that s_2 uses, i.e. $L(s_1) \subseteq L(s_2)$, and $U_{s_1}(x_{s_1}) < u_{s_1}^{\max}$, then $U_{s_1}(x_{s_1}) \geqslant U_{s_2}(x_{s_2})$.

It is a well-known property of the concept of proportional fairness that flows traversing a route receive a lower share of available resources than flows traversing a part of this route provided all utilities are equal. The rationale behind this is that flows using less resources should be favored to increase the total utility. Transferring this observation to utility proportional fairness, we get a similar result. Flows traversing several links receive less utility compared to shorter flows, provided a common transformation function is used. If this feature is undesirable, since the path a flow takes is chosen by the routing protocol and beyond the reach of the single user, the second order utilities can be modified to compensate this effect. We show that an appropriate choice of the transformation functions $f_s(\cdot)$ will assure a utility max–min bandwidth allocation in equilibrium.

Theorem 4. *Suppose all users have the same parameter dependent transformation function $f_s(q_s, \kappa) = q_s^{-\frac{1}{\kappa}}, s \in S$, $\kappa > 0$. The second order utilities $F_s(x_s, \kappa), s \in S$ are defined by (8). Let the sequence of rate vectors $\vec{x}(\kappa) = (x_s(\kappa) \in X_s, s \in S)$ be utility proportional fair. Then $\vec{x}(\kappa)$ approaches the utility max–min fair rate allocation as $\kappa \to \infty$.*

The proof of this theorem can be found in Appendix B.

4. Efficiency of utility fair resource allocations

Utility fair resource allocations ensure that an application specific performance measure, that is, the utility of received bandwidth, is shared fairly among sources. Our primary goal was to achieve this fairness requirement through distributed and stable congestion control algorithms. From the perspective of a system designer, however, one might be interested in quantifying how much utility is lost due to a fair allocation of utilities.

A degradation of system utility caused by different resource allocation goals have been analyzed before in many variants. A prominent example is the selfish behavior of users, who construct a solution such that their individual utility rather than social welfare is maximized, see Johari and Tsitsiklis [29] and the book by Roughgarden [30]. These works study the price of anarchy, which quantifies the sub-optimality of Nash equilibria compared to system optimal solutions. We are aware of few works that study the efficiency of suboptimal but fair solutions. Tang et al. [28,31] study the efficiency of heterogeneous transport protocols with respect to optimal solutions maximizing utility. They characterize the tradeoff between bandwidth fairness and throughput. None of these works, however, investigate the efficiency of utility fair solutions. Note that in contrast to the above mentioned works, the system optimum (solution of problem (1)) in our case is the solution to a non-convex program, which is known to be NP-hard in general. It is widely accepted that non-convex programs, unless they exhibit certain structures, are intractable in the sense that it is unlikely (unless $P = NP$) to find even efficient approximation algorithms, see Vavasis [32] and Bellare and Rogaway [33] for a survey on the complexity of non-convex programming.

We study in this section the efficiency of utility fair resource allocations with respect to maximum system utility as defined in (1). In the following, we refer to an optimal solution of (1) as the *system optimum*. For a given tuple $I = (N, R)$, vector of utility functions $\vec{U}$, and vector of transformation functions $\vec{f}$, we call the triple $(\vec{U}, \vec{f}, I)$ an *instance* of a utility proportional fair resource allocation. Recall that I characterizes the network, capacities, and the routes for a given set of sources.

Definition 4. We are given an *instance* $(\vec{U}, \vec{f}, I)$. Let $\vec{x}^*$ be the system optimal rate vector and $\vec{x}$ be the vector of utility

[1] A rate vector $\vec{x}$ is called Pareto efficient with respect to received rates (utilities) if no other feasible rate vector $\vec{y}$ exists such that $y_i > x_i (U_i(y_i) > U_i(x_i))$ for some $i \in S$ and $y_j \geqslant x_j (U_j(y_j) \geqslant U_j(x_j))$ for all $j \in S/\{i\}$.

proportional fair sending rates, see Definition 3. We define the efficiency of $\vec{x}$ with respect to $\vec{x}^*$ for $(\vec{U},\vec{f},I)$ as

$$\rho(\vec{U},\vec{f},I) := \frac{\sum_{s\in S} U_s(x_s)}{\sum_{s\in S} U_s(x_s^*)}.$$

Furthermore, we define the efficiency for given classes $\mathscr{U},\mathscr{F}$, and $\mathscr{I}$ as

$$\rho(\mathscr{U},\mathscr{F},\mathscr{I}) := \inf_{\vec{U}\in\mathscr{U}} \inf_{\vec{f}\in\mathscr{F}} \inf_{I\in\mathscr{I}} \rho(\vec{U},\vec{f},I).$$

Note that the values $\rho(\vec{U},\vec{f},I)$ and $\rho(\mathscr{U},\mathscr{F},\mathscr{I})$ are well defined as the solutions $\vec{x}$ and $\vec{x}^*$ to problems (10) and (1) exist and $\sum_{s\in S} U_s(x_s^*) > 0$. The case $\sum_{s\in S} U_s(x_s^*) = 0$ can be excluded since it implies $\vec{x} = 0$ for every feasible rate vector. By definition, we have $\rho(\vec{U},\vec{f},I) \in [0,1]$, where $\rho(\vec{U},\vec{f},I) = 1$ corresponds to full efficiency and $\rho(\vec{U},\vec{f},I) = 0$ to zero efficiency. We define the efficiency loss (or *price of fairness*) for given classes $\mathscr{U},\mathscr{F}$, and $\mathscr{I}$ by $1/\rho(\mathscr{U},\mathscr{F},\mathscr{I})$ if $\rho(\mathscr{U},\mathscr{F},\mathscr{I}) > 0$ and ∞, otherwise.

We first show that the price of fairness can be arbitrarily large for unrestricted classes $\mathscr{U}$ and $\mathscr{F}$.

Proposition 1. *For general classes $\mathscr{U}$ and $\mathscr{F}$ satisfying Assumptions 1 and 2 the price of fairness is unbounded, i.e. $\rho(\mathscr{U},\mathscr{F},\mathscr{I}) = 0$, even for single link networks.*

Proof. Consider an instance with a single link of capacity c and two linear utility functions $U_1(x_1) = \frac{1}{n}x_1, n \in \mathbb{N}$ and $U_2(x_2) = x_2$. Then, the system optimal solution is $\vec{x}^* = (0,c)$ with objective value c. Since we consider a single link, Theorem 3 implies that a utility proportional fair solution is also utility max–min fair. The corresponding rate vector is given by $\vec{x} = \left(\frac{c}{1/n+1}, \frac{c/n}{1/n+1}\right)$ with total utility $\frac{2c/n}{1/n+1}$. Letting n tend to infinity we have $\left(\frac{2c/n}{1/n+1}\right)/c \to 0$. Defining $\mathscr{U} = \{az : a \in [1/n,1], n \in \mathbb{N}\}$ we have $\inf_{U\in\mathscr{U}} \rho(\mathscr{U},\mathscr{F},\mathscr{I}) = 0$ proving the claim. $\square$

4.1. Efficiency loss for restricted utility and transformation functions

If the classes $\mathscr{U}$ and $\mathscr{F}$ are restricted, however, we can derive bounds on the efficiency loss of a utility proportional fair resource allocation. In the following, we present a generic technique to obtain a lower bound on the efficiency, which depends on the classes $\mathscr{U}$ and $\mathscr{F}$. For the tuple $I = (N,R)$, we define by $c(I) := \max\{c_l, l \in L\} > 0$ the maximum of all link capacities of the given network $N = (V,L,c)$. The value $c(\mathscr{I}) > 0$ is defined as the maximum $c(I), I \in \mathscr{I}$.

To this end, for a fixed utility function $U(x)$, transformation function $f(z)$, and capacity $c(I) > 0$, we define the following real-valued parameter:

$$\omega(U,f,c(I)) = \inf_{0\leqslant x,x^*\leqslant c(I)} \frac{U(x) + f^{-1}(U(x))(x^*-x)}{U(x^*)}. \tag{13}$$

Furthermore, we define

$$\omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) = \inf_{U\in\mathscr{U}} \inf_{f\in\mathscr{F}} \inf_{I\in\mathscr{I}} \omega(U,f,c(I)).$$

Before we present our main result in this section, we briefly comment on the above construction. In contrast to $\rho(\vec{U},\vec{f},I)$ the value $\omega(U,f,c(I))$ is defined for a single utility and transformation function, respectively. Furthermore, we restricted the feasible rates in the calculation of $\omega(U,f,c(I))$ by $c(I)$. This is certainly a rough restriction, as the equation $x = c(I)$ cannot hold for all rates simultaneously. The striking advantage of this relaxation, however, is that we can reduce the calculation of $\omega(U,f,c(I))$ to minimizing a real-valued functions on a simple interval. Furthermore, the objective becomes separable in the rates, utility, and transformation functions.

It is easy to verify that $\omega(U,f,c(I)) \in (-\infty,1]$. The following theorem provides a lower bound on the price of fairness.

Theorem 5. *For given classes $\mathscr{U},\mathscr{F}$, and $\mathscr{I}$, the efficiency $\rho(\mathscr{U},\mathscr{F},\mathscr{I})$ of a utility proportional fair rate vector is bounded from below by $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$.*

Proof. We are given an arbitrary instance $(\vec{U},\vec{f},I)$ with $U \in \mathscr{U}, F \in \mathscr{F}$, and $I \in \mathscr{I}$. Let $\vec{x}^*$ denote the system optimal rate vector. We start with the system utility of the utility proportional fair rate vector $\vec{x}$:

$$\begin{aligned}\sum_{s\in S} U_s(x_s) &\geqslant \sum_{s\in S} U_s(x_s) + \sum_{s\in S} f_s^{-1}(U_s(x_s))(x_s^* - x_s)\\ &= \sum_{s\in S} \left(U_s(x_s) + f_s^{-1}(U_s(x_s))(x_s^* - x_s)\right)\\ &\geqslant \sum_{s\in S} \omega(U_s,f_s,c(I))U_s(x_s^*)\\ &\geqslant \omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) \sum_{s\in S} U_s(x_s^*).\end{aligned}$$

The first inequality follows from (12), where we set $y_s = x_s^*$. The second and third inequality follow from the definition of $\omega(U,f,c(I))$ and $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$, respectively. $\square$

Whenever $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) \leqslant 0$, Theorem 5 does not yield an approximation guarantee. In fact, as we have seen from Proposition 1, the value $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$ is 0 (or even smaller) for general classes $\mathscr{U}$ and $\mathscr{F}$. If the classes $\mathscr{U},\mathscr{F}$, and $\mathscr{I}$ are restricted, however, it is possible to derive an approximation guarantee of $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) > 0$.

Theorem 5 can be used with respect to three aspects:

(1) for a given instance $(\vec{U},\vec{f},I)$, the question of deriving an upper bound on the efficiency loss boils down to minimizing a single real-valued function $\omega(U_s,f_s,c(I))$ for all $s \in S$;
(2) the behavior of $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$ can be used to *design* the set of utility and transformation functions $\mathscr{U}$ and $\mathscr{F}$ so as to achieve a certain level of efficiency;
(3) any set $\mathscr{F}$ having $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) > 0$ induces a polynomial-time approximation algorithm for the NP-hard problem (1) with approximation guarantee $1/\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$.

Before we explicitly calculate bounds on the efficiency loss for restricted classes of utility functions, we comment on the limitations of our approach.

Remark 3. Bellare and Rogaway [33] showed that even for simple quadratic programming problems with linear constraints, there exists no polynomial-time μ-approximation algorithm and $\mu \in (0,1)$, unless $P = NP$. As a utility proportional fair solution can be computed in polynomial time (see the convex problem (10)) it is unlikely that one can prove an approximation guarantee of $\mu \in (0,1)$ even when the objective function in (1) is restricted to quadratic polynomials.

4.2. Linear, concave, and bounded non-concave utility functions

In light of Remark 3, we will make several simplifying assumptions. We will analyze the efficiency loss of utility proportional fair rate vectors, when utility functions are linear, concave, or contained in a sector whose boundaries are given by linear functions. These cases are of interest, when the number of elastic traffic sources is large and only a few inelastic applications are present, or, when the nonlinearity of utility functions is bounded. We will quantify in the following, to which extend our fairness criterion impacts overall utility.

Corollary 1 (of Theorem 5). *Assume $\mathscr{F} = \{f : f(z) = \frac{1}{z} - 1\}$ represents the class of transformation functions. The set of utility functions is given by $\mathscr{U} = \{U : U(x) = ax, a \in [1,b], b \in [1,\infty)\}$ and let $c(\mathscr{I}) \geqslant 2$. Then, the efficiency is at least*

$$\rho(\mathscr{F},\mathscr{U},c(\mathscr{I})) \geqslant \frac{1}{bc(\mathscr{I})}.$$

Proof. We are given an arbitrary instance $(\vec{U},\vec{f},I)$ with $U \in \mathscr{U}, F \in \mathscr{F}$, and $I \in \mathscr{I}$. Since the transformation function is given by $f(z) = \frac{1}{z} - 1$, we have for its inverse $f^{-1}(u) = \frac{1}{u+1}$. Then, applying the definition of $\omega(U,f,c(I))$ and $\omega(\mathscr{U},\mathscr{F},c(\mathscr{I}))$ yields

$$\omega(\mathscr{U},\mathscr{F},c(\mathscr{I})) = \inf_{\substack{0 \leqslant x,y \leqslant c(I)\\ 1 \leqslant a \leqslant b}} \frac{ax + \frac{1}{ax+1}(y-x)}{ay} = \inf_{\substack{0 \leqslant x,y \leqslant c(I)\\ 1 \leqslant a \leqslant b}} \frac{x}{y} + \frac{y-x}{(ax+1)ay}.$$

Let us define the function $g(a,x,y) = \frac{x}{y} + \frac{y-x}{(ax+1)ay}$, which we want to minimize. To solve the minimum, we first evaluate $\frac{\partial g(a,x,y)}{\partial y} = -\frac{x(a^2x+a-1)}{(ax+1)ay^2}$. Since $a \geqslant 1$, we have $\frac{\partial g(a,x,y)}{\partial y} \leqslant 0$, implying $y^* = c(I)$. The function $g(a,x,c(I))$ is given by

$$g(a,x,c(I)) = \frac{x}{c(I)} + \frac{c(I)-x}{(ax+1)ac(I)}.$$

This function is strictly convex with respect to x, having its global minimum at

$$x^* = \left[\frac{-ac(I) + \sqrt{ac(I) + a^2c(I)}}{a^2}\right]^+.$$

Using $c(\mathscr{I}) \geqslant 2$ implies $c(I) \geqslant \sqrt{2c(I)}$. Thus, the positive projection is active, i.e., $x^* = 0$. Inserting this solution we arrive at $g(a,0,c(I)) = \frac{1}{ac(I)}$. This function is strictly decreasing with respect to a, hence, the global minimum is given by $g(b,0,c(I)) = \frac{1}{bc(I)}$ proving the first claim. The assumption $c(\mathscr{I}) \geqslant 2$ was made for simplifying the analysis.

Note that the lower bound in Corollary 1 holds for arbitrary networks and number of users. We do not know if the lower bound of $\frac{1}{bc(\mathscr{I})}$ is tight. We now present a tight result for single link networks, where we allow for non-concave utility functions contained in a sector, which is enclosed by linear utility functions, see Fig. 2 for an illustration.

Proposition 2. *The set of utility functions is given by*

$$\mathscr{U} = \{U : U(x) \in [ax, bx],\ \textit{for all}\quad x \in [0, c(\mathscr{I})],\ a \in [0,b],\ \textit{and}\ b \in (0,\infty)\}.$$

Let $\mathscr{I}$ be a class of single link networks. Then, $\rho(\mathscr{F},\mathscr{U},c(\mathscr{I})) \geqslant \frac{a}{b}$. Furthermore, this bound is tight.

Proof. Let $\vec{x}^*$ denote the system optimum and let $\vec{x}$ denote the utility proportional fair rate vector for an instance $(\vec{U},\vec{f},I)$ satisfying the above conditions. Without loss of generality, we can assume $x_s^{\max} \geqslant c(I)$ for all $s \in S$. First we derive a lower bound on the overall utility of the system optimum $\vec{x}^*$. Using $U_s(x_s^*) \leqslant bx_s^*$ for all $x_s^* \in [0,c(I)]$, $s \in S$, we have $\sum_{s\in S} U_s(x_s^*) \leqslant bc(I)$. Conversely, we know that $U_s(x_s) \geqslant ax_s$ for all $x_s \in [0,c(I)]$, $s \in S$. This together with $x_s^{\max} \geqslant c(I)$ implies $\sum_{s\in S} U_s(x_s) \geqslant ac(I)$ proving the lower bound.

For proving the upper bound, we consider a single link of capacity $c(I)$ and $n+1$ sources having utility functions $U(x_i) = ax_i, i = 1,\dots,n$ and $U(x_{n+1}) = bx_{n+1}$. We assume that every user has the same transformation function. Thus, Theorem 3 implies that the resulting utility proportional fair rate allocation is utility max–min fair. Assuming that $x_s^{\max} \geqslant c(I)$ for all $s \in S$, we know that every user receives the same utility. The utility max–min fair rate vector is given by $\vec{x} = \left(\frac{c(I)}{a/b+n},\dots,\frac{c(I)}{a/b+n},\frac{ac(I)}{a+bn}\right)$. The total utility evaluates to $\frac{nc(I)}{a/b+n} + \frac{bac(I)}{a+bn} \to ac(I)$ for $n \to \infty$. The optimal rate vector is $\vec{x}^* = (0,\dots,0,c(I))$ with total utility $bc(I)$ proving the second claim. □

Corollary 2 (of Theorem 5). *Assume the classes $\mathscr{U}$, $\mathscr{F}$, and $\mathscr{I}$ satisfy*

$$\lambda U'(x) \leqslant f^{-1}(U(x)) \leqslant U'(x)$$

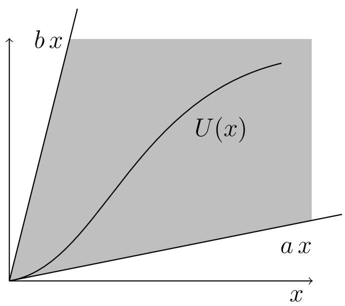


Fig. 2. Illustration of the sector of feasible utility functions in Proposition 2.

for all $x \leqslant c(\mathscr{I}), 0 \leqslant \lambda \leqslant 1$, for all $U \in \mathscr{U}, f \in \mathscr{F}, I \in \mathscr{I}$. If additionally $\mathscr{U}$ is a class of concave utility functions, then, the efficiency is bounded from below by λ.

Proof. We prove the theorem by analyzing the parameter $\omega(U, f, c(I))$ for $U \in \mathscr{U}, f \in \mathscr{F}$, and $I \in \mathscr{I}$.

$$\omega(U, f, c(I)) = \inf_{0 \leqslant x, y \leqslant c(I)} \frac{U(x) + f^{-1}(U(x))(y-x)}{U(y)}.$$

First, assume $x \geqslant y$. In this case, using $f^{-1}(U(x)) \leqslant U'(x)$ we have

$$\frac{U(x) + f^{-1}(U(x))(y-x)}{U(y)} \geqslant \frac{U(x) + U'(x)(y-x)}{U(y)}.$$

Concavity of $U(\cdot)$ implies $U(x) + U'(x)(y-x) \geqslant U(y)$, hence $\omega(U, f, c(I)) \geqslant 1$. Now, assuming $x < y$ implies:

$$\frac{U(x) + f^{-1}(U(x))(y-x)}{U(y)} \geqslant \frac{U(x) + \lambda U'(x)(y-x)}{U(y)}.$$

Then, using again concavity of $U(\cdot)$ and $0 \leqslant \lambda \leqslant 1$ we have

$$\frac{U(x) + \lambda U'(x)(y-x)}{U(y)} \geqslant \frac{U(x) + \lambda(U(y) - U(x))}{U(y)} \geqslant \lambda. \qquad \square$$

Fig. 3 illustrates the feasible set $\mathscr{F}$ of transformation functions such that the function $f^{-1}(U(x))$ stays in the corridor between $U'(x)$ and $\lambda U'(x)$ for some $\lambda \in (0, 1)$.

In the following, we present conditions implying full efficiency of a utility proportional fair rate vector. For the next result, we require a condition on the transformation functions, which in turn implies the standard necessary (and sufficient) optimality condition for the concave utility maximization problem.

Proposition 3. *We are given an instance $(\vec{U}, \vec{f}, I)$, where $\vec{U}$ is a vector of strictly concave utility functions. The utility proportional fair rate vector $\vec{x}$ for the instance $(\vec{U}, \vec{f}, I)$ has $\rho(\vec{U}, \vec{f}, I) = 1$ if*

$$f_s^{-1}(U_s(x_s)) = U_s'(x_s), \text{ for all } s \in S. \qquad (14)$$

Proof. Suppose that $\vec{x}$ is utility proportional fair. Thus, by Definition 3 the rate vector $\vec{x}$ satisfies:

$$\sum_{s \in S} f_s^{-1}(U_s(x_s))(y_s - x_s) \leqslant 0,$$

for any feasible rate vector $\vec{y}$. Using assumption (14), we have

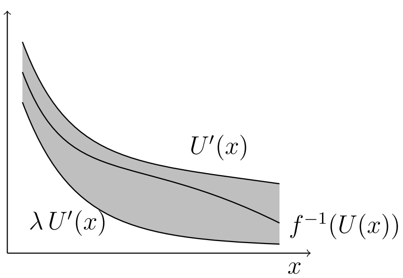


Fig. 3. Illustration of the classes $\mathscr{F}$ of feasible transformation functions for a given utility function $U(x)$. The gray shaded area corresponds to the feasible space for $f^{-1}(U(x))$.

$$\sum_{s \in S} U_s'(x_s)(y_s - x_s) \leqslant 0,$$

for any feasible $\vec{y}$, proving that $\vec{x}$ is an optimal solution to (1). $\square$

Alternatively, it is possible to prove this proposition by analyzing the parameter $\omega(\mathscr{U}, \mathscr{F}, c(\mathscr{I}))$. We present a simple example for this result. Consider the utility function $U(x) = \sqrt{x}$ and consider the transformation function $f(z) = 1/(2z)$. Then, we have $U'(x) = 1/(2\sqrt{x}) = f^{-1}(U(x))$. Thus, if all utility functions and transformation functions are of the above form, a utility proportional fair rate vector is fully efficient.

5. Fair utility functions

Traditionally, a bandwidth allocation is considered fair if flows get (approximately) equal shares of the available bandwidth, i.e. $x_i \approx x_j, i, j \in S$. In utility fair networks, this relation holds for utilities, i.e. $U_i(x_i) \approx U_j(x_j)$. Equalizing utility does not imply that bandwidth is shared equally. In fact, inelastic applications, such as VoIP traffic, are not able to adapt the sending rate with arbitrary granulation. It is reasonable, however, to investigate bandwidth fairness over a time period assuming that the state of congestion varies over time. In this regard, we introduce the notion of long-term bandwidth fairness in utility fair networks. We model the change of congestion over time by a probability distribution of the available utility $f(q)$. To assess long-term bandwidth fairness of utility functions we say that two utility functions $U_1(x_1)$ and $U_2(x_2)$ are bandwidth fair with respect to a probability distribution of the available utility $f_{1,2}(q)$, if the *expected* bandwidth allocations are equal, i.e. $E[x_1] = E[x_2]$. In the following, we assume that the available utility is uniformly distributed. Then, we can define the following fairness measure to compare different utility functions:

Definition 5. The fairness measure $\delta_s(U_s, X)$ on the interval $X = [x_{\min}, x_{\max}]$ is defined as

$$\delta_s(U_s, X) := \int_{X \cap X_s} U_s(x_s) dx_s + \int_{\min\{x_{\max}, x_s^{\max}\}}^{x_{\max}} u_{\max} dx_s. \qquad (15)$$

If $\delta_s(U_s, X) = r, r > 0$, then, $U_s(x_s)$ is said to be r-fair in X.

See Fig. 4 for a graphical depiction of the terms in this definition. This measure implies a fairness metric for utility functions: an application/user s' with fairness measure $\delta_{s'}(U_{s'}, X) < \delta_s(U_s, X)$ for its utility function will get on average more bandwidth on the interval X than user s. In the following Theorem, we formalize this relation.

Theorem 5. *Suppose, two users $s = 1, 2$ use the path L_p through the network. and their utility functions are r_1 and r_2-fair on the interval $X = [x_{\min}, x_{\max}]$. Assume the path price q on L_p varies in the interval $[q_{\min}, q_{\max}]$, so that the available utility $f_s(q), s = 1, 2$, is uniformly distributed on the interval $[u_{\min}, u_{\max}]$. Then, the following conditions hold:*

(1) If $r_1 = r_2$, then, $E[x_1] = E[x_2]$.
(2) If $r_1 \leqslant r_2$, then, we have $E[x_1] \geqslant E[x_2]$.

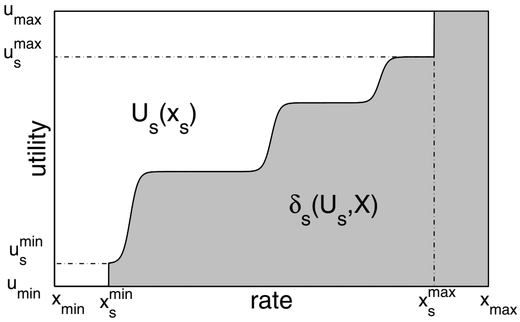


Fig. 4. Fairness measure $\delta_s(U_s, X)$.

Proof. Since $f_s(q_s), s = 1, 2$, are uniformly distributed on $[u_{\min}, u_{\max}]$, the corresponding probability density functions are given by $\frac{1}{u_{\max}-u_{\min}}$. Using (6), the expected bandwidth share for user 1 on the interval $[u_{\min}, u_{\max}]$ is given as

$$E[x_1] = \int_{u_{\min}}^{u_{\max}} \frac{[U_1^{-1}(\tau)]_{x_1^{\min}}^{x_1^{\max}}}{u_{\max} - u_{\min}} \mathrm{d}\tau,$$

where τ represents the realization of the available utility. Due to symmetry, we have:

$$\begin{aligned} E[x_1] &= \int_{u_{\min}}^{u_{\max}} \frac{\left[U_1^{-1}(\tau)\right]_{x_1^{\min}}^{x_1^{\max}}}{u_{\max} - u_{\min}} \mathrm{d}\tau, \\ &= (x_{\max} - x_{\min})(u_{\max} - u_{\min}) - r_1, \\ &= (x_{\max} - x_{\min})(u_{\max} - u_{\min}) - r_2, \\ &= E[x_2]. \end{aligned}$$

Using $r_1 \leqslant r_2$, we immediately get $E[x_1] \geqslant E[x_2]$.

The above Theorem allows us to give an alternative definition of traditional *TCP-friendliness* in the context of utility fair networks. As shown in [3], it is possible to reverse engineer the underlying utility functions of TCP. In our alternative definition, we consider a real-time application to be TCP-friendly over a certain bandwidth interval, if the corresponding utility function has the same fairness measure as the underlying TCP utility function. The interpretation is different from the original TCP-friendliness paradigm though. Due to its inelasticity, a real-time flow may not be able to adapt the sending rate with arbitrary granularity (e.g. layered multimedia) and behave as aggressive as TCP would. The alternative definition rather indicates that an application with a TCP-fair utility function will get on *average* (light loaded network versus heavily loaded network) as much bandwidth as a TCP flow would. How this fairness metric can be used as the basis for a pricing framework is shown in [23].

6. Simulation results

To demonstrate the effectiveness of our utility-based congestion control approach we performed simulations using the *NS-2* network simulator [34] emphasizing on adaptation of real-time flows, prioritization and utility proportional fair bandwidth allocation.

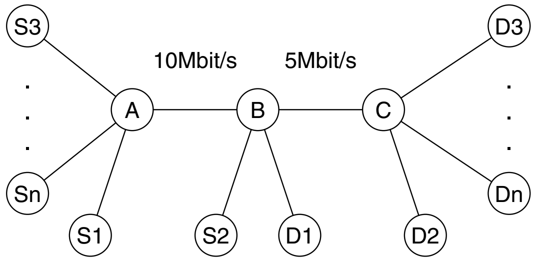


Fig. 5. Network topology.

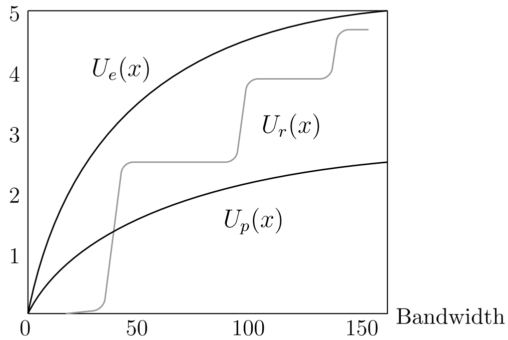


Fig. 6. Utility functions used for the simulations.

Our senders are rate-based, i.e., packets are sent at intervals of $1/x_s$ seconds. We limit the changes of x_s to not increase by more than 1 and not decrease by more than half in one round-trip-time to smooth sender behavior on startup and when adapting to changing network load. Link prices are accumulated in a double precision floating point field in the packet headers that receivers return to the senders in acknowledgment packets. Data packets have a fixed size of 1500 bytes and receivers acknowledge every data packet immediately.

Fig. 5 shows the network topology and Fig. 6 the bandwidth utility functions used in the simulations. There are two elastic utility functions, U_e and prioritized U_p[2], and one real-time utility function U_r that models a video streaming application with three supported coding layers: 512 kbit/s ($c_{r,1} = 43$ pps), 1024 kbit/s ($c_{r,2} = 84$ pps), and 1.5 Mbit/s ($c_{r,3} = 125$ pps).

All senders use the same transformation function $f_s(q_s) = q_s^{-\frac{1}{k}}$. The utility functions are given by

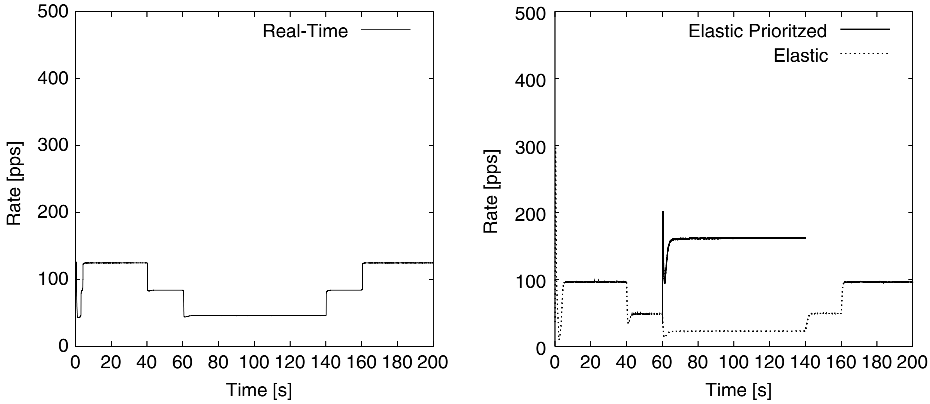
$$U_r(x) = \begin{cases} \frac{2}{1+\exp(-(x-43))}, 0 \leqslant x < 63.5, \\ \frac{1.5}{1+\exp(-(x-84))} + 2, 63.5 \leqslant x < 104.5, \\ \frac{1}{1+\exp(-(x-125))} + 2 + 1.5, 104.5 \leqslant x \leqslant 130, \end{cases}$$

$$U_e(x) = 0.4\sqrt{x},$$

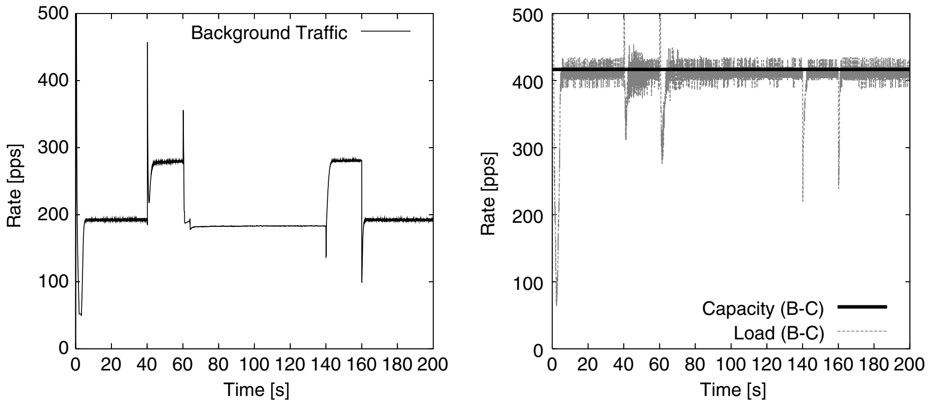
$$U_p(x) = 0.15\sqrt{x}.$$

The results of four simulation setups are shown in Fig. 7a and 7b.

[2] For a given utility value, a prioritized flow receives a larger bandwidth share than competing non-prioritized flows.



(a) Adaptation and prioritization of flows



(b) Adaptation and prioritization of flows

Fig. 7. Simulation results.

Our first simulation focuses on adaptation of a real-time flow when network load changes and on prioritization of flows. The four graphs in Fig. 7a and 7b show the sending rate of one real-time flow (using U_r), the rates of two elastic flows (prioritized U_p and non-prioritized U_e) and the aggregated rates of additional background traffic (up to four elastic and two real-time flows), which together yield the total load shown in the fourth graph. Starting with the highest encoding rate (125 pps), the real-time flow switches to lower rates as more flows start after 40 and 60 s. Flows terminate after 140 and 160 s which allows the real-time flow to switch back to higher encoding rates. The prioritized elastic flow starts after 60 and ends after 140 s and receives a significant higher bandwidth share, although it receives the same utility as the other flows.

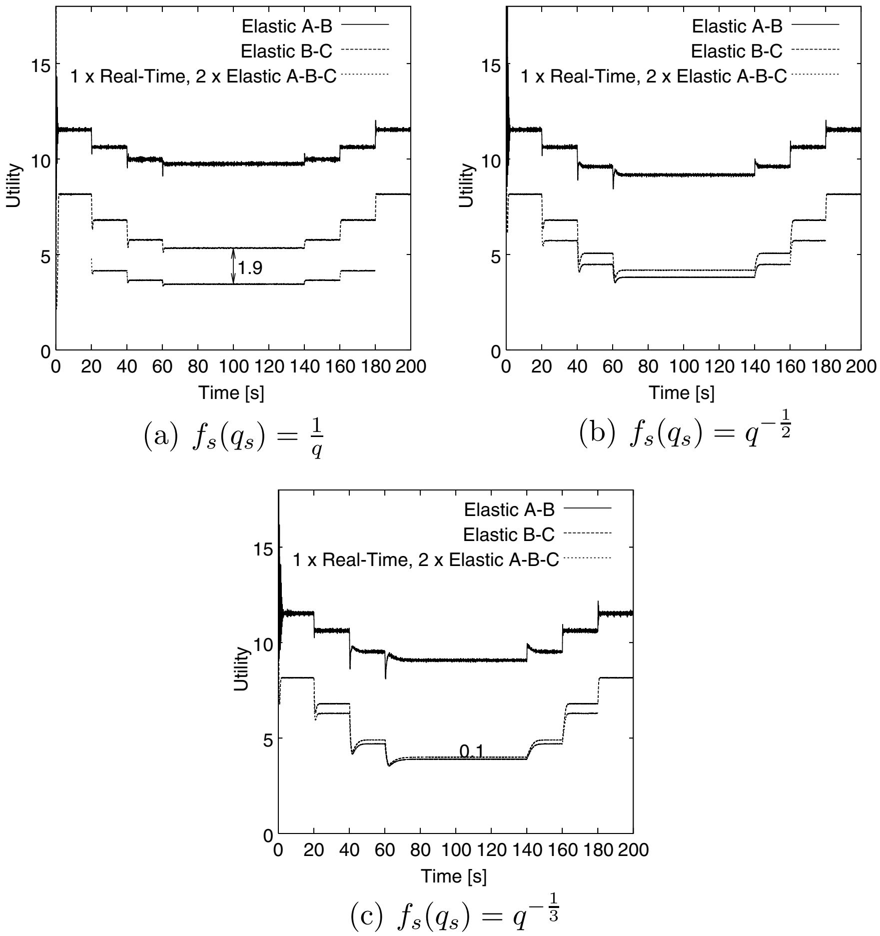
In this setup, there are no short flows using only the A–B or B–C links and all senders have $\kappa = 1$.

The remaining three Fig. 8a, 8b and 8c show the effect of an increasing κ for $\kappa = 1$, 2 and 3 on the resulting bandwidth shares. Here, two short flows using the A–B and B–C links, respectively, compete with up to three long flows using the whole A–B–C path. As can be seen, the difference in received utility between short and long flows competing for the same bottleneck link (B–C) decreases for increasing values of κ. Thus, these results demonstrate the convergence towards a utility max–min fair equilibrium for increasing values of κ as stated in Theorem 4.

7. Conclusion

We have obtained decentralized congestion control laws at links and sources, which are locally stable and provide a utility proportional fair resource allocation in equilibrium. Our fairness criterion ensures that bandwidth utility values of users (applications), rather than bit rates, are proportional fair in equilibrium. In a limiting case of our model, we incorporated utility max–min fairness for all users sharing the network. We further investigated the efficiency of utility fair resource allocation with respect to maximum total utility. Finally, we developed a fairness measure mapping the specific shape of a utility function to a fairness value describing its expected bandwidth consumption.

As an important open problem, we see the development of globally stable congestion control algorithms for time-varying delays in combination with non-concave utility functions. We refer to recent work of Harks and Miller [19] as a first promising step towards this goal.



(a) $f_s(q_s) = \frac{1}{q}$ (b) $f_s(q_s) = q^{-\frac{1}{2}}$ (c) $f_s(q_s) = q^{-\frac{1}{3}}$

Fig. 8. Simulation results.

Appendix A

A.1. Proof of Theorem 3

We prove that it is impossible to strictly increase the utility $U_s(y_s) > U_s(x_s)$ for a user $s \in S_{L_p}$ without strictly decreasing the utility $U_k(y_k) < U_k(x_k)$ for some $k \in S_{L_p}$, which is already smaller than $U_s(x_s)$, that is, $U_k(x_k) \leqslant U_s(x_s)$. Using Definition 1 this would prove the claim.

First note that we cannot strictly increase the utility of a user without strictly decreasing the utility of at least another user. To see this recall that the rate vector $\vec{x}$ maximizes second order utility, which is a strictly increasing function of the utility. To prove the theorem we have to show that in order to strictly increase the utility of a user we have to strictly decrease the utility of at least another user, *whose utility is already smaller*.

Before we prove this, observe the following useful facts:

$$U_s(x_s) = u_s^{\min} \Rightarrow x_s = 0, \tag{16}$$
$$U_s(x_s) = u_s^{\max} \Rightarrow x_s = x_s^{\max}. \tag{17}$$

Both facts follow from Assumption 1 and since utility functions are strictly increasing.

Using the equilibrium structure (5) for users $s \in S_{L_p}$, we can partition the set S_{L_p} into the sets

$$Q_1 := \{s \in S_{L_p} | U_s(x_s) = u_s^{\max} < f(q_p)\},$$
$$Q_2 := \{s \in S_{L_p} | U_s(x_s) = f(q_p)\},$$
$$Q_3 := \{s \in S_{L_p} | U_s(x_s) = u_s^{\min} > f(q_p)\}.$$

Then, we can see that for any users $s_1 \in Q_1, s_2 \in Q_2, s_3 \in Q_3$ the following relation is valid:

$$U_{s_1}(x_{s_1}) < U_{s_2}(x_{s_2}) < U_{s_3}(x_{s_3}). \tag{18}$$

We will show that (i.a) a strict increase of the utility of a user in Q_3 results in a strict decrease of the utility of at least one user in $Q_1 \cup Q_2$; (i.b) a strict increase of the utility of a user in Q_2 results in a strict decrease of the utility of at least one user in $Q_1 \cup Q_2$. We do not have to consider the case of strictly increasing the utility of a user $s \in Q_1$ because $U_s(x_s) = u_s^{\max}$ for $s \in Q_1$.

Before we prove (i.a) and (i.b), we show how it implies the claim. Condition (i.a) states that in order to strictly increase the utility of a user in Q_3, we have to strictly decrease the utility of a user in $Q_1 \cup Q_2$. Using (18) we know that the utility we have to strictly decrease is already smaller. Condition (i.b) states that in order to strictly increase the utility of a user in Q_2, we have to strictly decrease the utility of a user in $Q_1 \cup Q_2$. If we have to strictly decrease the utility of a user in Q_1, then we can apply the same argument, see (18). If we have to strictly decrease the utility of a user in Q_2, then, knowing that the utility for all users in Q_2 is equal, we have to strictly decrease the utility of another user in Q_2, whose utility is already smaller.

(i.a) Suppose we strictly increase $U_s(x_s)$ to $U_s(y_s) > U_s(x_s)$ for some $s \in Q_3$. Since utility functions are strictly increasing and (16) implies $x_s = 0$ for all $s \in Q_3$ we have to strictly decrease at least one of the rates $x_s, s \in Q_1 \cup Q_2$. If we strictly decrease the rate x_s for some $s \in Q_1$ ((17) implies $x_s = x_s^{\max}$ for $s \in Q_1$), we strictly decrease the utility $U_s(x_s)$ because utility functions are strictly increasing in $X_s = [0, x_s^{\max}]$. We are left with the case that we have to strictly decrease at least one of the users in Q_2. Note that for a user in Q_2 we either have $U_s(x_s) = u_s^{\max}, U_s(x_s) = u_s^{\min}$, or $u_s^{\min} < U_s(x_s) = f(q_p) < u_s^{\max}$. The first case $U_s(x_s) = u_s^{\max}$ can be treated as before. The second case $U_s(x_s) = u_s^{\min}$ can be omitted, since it implies $x_s = 0$, thus it is impossible to decrease this rate. In the third case, we again use that utility functions are strictly increasing in $X_s = [0, x_s^{\max}]$ implying that we have to strictly decrease the utility of this user.

(i.b) Suppose we strictly increase $U_s(x_s)$ to $U_s(y_s) > U_s(x_s)$ for some $s \in Q_2$. This implies that we have to strictly decrease the rate of at least one user in $Q_1 \cup Q_2$ (note that for users in Q_3 we have $x_s = 0$). The strict monotonicity of the utility functions implies that we have to strictly reduce the utilities of at least one of the users in $Q_1 \cup Q_2$.

To (ii): Assume $q_{s_1} \in Q_{s_1}, q_{s_2} \in Q_{s_2}$ and $q_{s_1} \leqslant q_{s_2}$ for sources s_1, s_2. Applying (5) to given q_{s_1}, q_{s_2}, we have $f(q_{s_1}) = U_{s_1}(x_{s_1}) \geqslant f(q_{s_2}) = U_{s_2}(x_{s_2})$ because of the monotonicity of $f(\cdot)$.

To (iii): From $L(s_1) \subseteq L(s_2)$ it follows, that $q_{s_1} \leqslant q_{s_2}$. Since the available utility $f(\cdot)$ is monotone decreasing in q_s and the bandwidth utility $U_{s_1}(x_{s_1}) < u_{s_1}^{\max}$ of user s_1 is not bounded by its maximum value, it follows, that $f(q_{s_1}) = U_{s_1}(x_{s_1}) \geqslant [f(q_{s_2})]_{u_{s_2}^{min}}^{u_{s_2}^{\max}} = U_{s_2}(x_{s_2})$. □

Appendix B

B.1. Proof of Theorem 4

Since all elements of the sequence $\vec{x}(\kappa)$ solve (10) subject to linear constraints, the sequence is bounded. Hence, we find a subsequence $\vec{x}(\kappa_p), p \in \mathbb{N}^+$, such that $\lim_{p\to\infty}\vec{x}(\kappa_p) = \vec{x}$. We show, that this limit point $\vec{x}$ is utility max–min fair. The uniqueness of the utility max–min fair rate vector $\vec{x}$ will ensure that every limit point of $\vec{x}(\kappa)$ is equal $\vec{x}$. This proves the convergence of $\vec{x}(\kappa)$ to $\vec{x}$. Since all users $s \in S$ use the same transformation function $f_s(q_s) = q_s^{-\frac{1}{\kappa}}, s \in S$, the second order utility and its derivative applied to the rate vector $\vec{x}(\kappa)$ have the following form:

$$F_s(x_s(\kappa)) = \int U_s(x_s(\kappa))^{-\kappa} \mathrm{d}x_s(\kappa) \text{ with } F'_s(x_s(\kappa)) = U_s(x_s(\kappa))^{-\kappa}, s \in S.$$

The proof now proceeds by contradiction and assuming that the limit point $\vec{x}$ is not utility max–min fair. Then, the idea is to construct a feasible rate vector $\vec{y}$, which receives higher second order utility than $\vec{x}$. This leads to a contradiction to the fact that $\vec{x}$ maximizes second order utility, thus, proving the claim.

Assume that the limit point $\vec{x} = (x_s \in X_s,\ s \in S)$ is not utility max–min fair. Then we can increase the bandwidth utility of a user j while decreasing the utilities of other users $k \in K$ which are larger than $U_j(x_j)$. More formal, there exists an index j, a set $K \subset S \setminus \{j\}$, and a feasible rate vector $\vec{y}$ with

$$y_s = \begin{cases} x_s, & \text{if } s \in S/\{K \cup \{j\}\}, \\ y_s, & \text{if } s \in K \cup \{j\}, \end{cases}$$

such that $U_j(y_j) > U_j(x_j)$ and $U_k(y_k) < U_k(x_k)$ for all $k \in K$, with $U_k(y_k) > U_j(x_j)$. We can omit the case $K = \emptyset$ since this would imply that we can strictly increase second order utility without decreasing any other rates. This contradicts the fact that $\vec{x}$ maximizes second order utility. Now, we choose κ_0 so large that for all elements of the subsequence $\vec{x}(\kappa_p)$ with $\kappa_p > \kappa_0$ the inequalities $U_j(y_j) > U_j(x_j(\kappa_p))$ and $U_k(y_k) < U_k(x_k(\kappa_p))$ with $U_k(y_k) > U_j(x_j(\kappa_p))$ for $k \in K$ hold. With the inequality $U_j(x_j(\kappa_p)) < U_k(x_k(\kappa_p)),\ k \in K$, we can choose $\kappa_1 > \kappa_0$ large enough such that

$$U_j(x_j(\kappa_p))^{-\kappa_p} > C \cdot U_k(x_k(\kappa_p))^{-\kappa_p} \tag{19}$$

holds for all $k \in K, \kappa_p > \kappa_1$, and $C > 0$ an arbitrary constant. Hence, there exists a κ_1 large enough that the following inequality holds:

$$U_j(x_j(\kappa_p))^{-\kappa_p} \underbrace{(y_j - x_j(\kappa_p))}_{>0} > \sum_{k\in K} \underbrace{(x_k(\kappa_p) - y_k)}_{>0} \max_{k\in K} U_k(x_k(\kappa_p))^{-\kappa_p}, \tag{20}$$

for all $\kappa_p > \kappa_1$. We evaluate the variational inequality (12) given in the definition of utility proportion fairness for the candidate rate vector $\vec{y}$ and $\kappa_p > \kappa_1$.

$$\begin{aligned}\sum_{s\in S} F'_s(x_s(\kappa_p))(y_s - x_s(\kappa_p)) &= \sum_{s\in S} U_s(x_s(\kappa_p))^{-\kappa_p}(y_s - x_s(\kappa_p)) \\ &= U_j(x_j(\kappa_p))^{-\kappa_p}(y_j - x_j(\kappa_p)) \\ &\quad + \sum_{k\in K} U_k(x_k(\kappa_p))^{-\kappa_p}(y_k - x_k(\kappa_p)) \\ &\geqslant U_j(x_j(\kappa_p))^{-\kappa_p}(y_j - x_j(\kappa_p)) \\ &\quad - \max_{k\in K} U_k(x_k(\kappa_p))^{-\kappa_p} \sum_{k\in K}(x_k(\kappa_p) - y_k) \\ &> 0, \text{using } (20).\end{aligned}$$

Hence, the variational inequality is not valid contradicting the utility proportional fairness property of $\vec{x}(\kappa_p)$. □

References

[1] S.H. Low, D.E. Lapsley, Optimization Flow Control I, IEEE/ACM Transaction on Networking 7 (6) (1999) 861–874.

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

[3] S. Low, F. Paganini, J.C. Doyle, Internet Congestion Control, IEEE Control Systems Magazine 22.

[4] J. Mo, J. Walrand, Fair end-to-end window-based congestion control, IEEE/ACM Transaction on Networking 8 (5) (2000) 556–567.

[5] S. Shenker, Fundamental design issues for the future internet, IEEE JSAC 13 (1995) 1176–1188.
[6] J.-W. Lee, R. Mazumdar, N.B. Shroff, Non-convex optimization and rate control for multi-class services in the internet, IEEE/ACM Transaction Network 13 (4) (2005) 827–840.
[7] S. Sarkar, L. Tassiulas, Fair allocation of utilities in multirate multicast networks: A framework for unifying diverse fairness objectives, IEEE Transaction on Automatic Control 47 (6) (2002) 931–944.
[8] Z. Cao, E. Zegura, Utility max–min: An application-oriented bandwidth allocation scheme, in: Proceedings of IEEE INFOCOM, 1999, pp. 793–801.
[9] R.F. Liao, T. Campbell, A utility-based approach for quantitative adaption in wireless packet networks, Wireless Networks 7 (5) (2001) 541–557.
[10] R. Srikant, The Mathematics of Internet Congestion Control, Birkhauser, Boston, 2003.
[11] F. Paganini, A global stability result in network flow control, Systems and Control Letters 46 (2002) 165–172.
[12] F. Paganini, Z. Wang, J. Doyle, S. Low, Congestion control for high performance, stability and fairness in general networks, IEEE/ACM Transaction on Networking 13 (1) (2005) 43–56.
[13] M. Chiang, S. Zhang, P. Hande, Distributed rate allocation for inelastic flows: optimization frameworks, optimality conditions, and optimal algorithms, in: Proceedings of IEEE INFOCOM, Miami, FL, 2005, pp. 2679–2690.
[14] M. Fazel, M. Chiang, Network utility maximization with nonconcave utilities using sum-of-squares method, in: Proceedings of 44th IEEE Conference Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05., 2005, pp. 1867–1874.
[15] T. Harks, Utility proportional fair resource allocation: an optimization oriented approach, in: Proceedings of QoS in Multiservice IP Networks, Springer, Catania, Italy, 2005, pp. 61–74.
[16] H.-W. Lee, S. Chong, A distributed utility max–min flow control algorithm, Computer Networks 50 (11) (2006) 1816–1830.
[17] J. Cho, S. Chong, Utility max–min flow control using slope-restricted utility functions, IEEE Transaction on Communications 55 (2007) 963–972.
[18] Y. Zhang, S.-R. Kang, D. Loguinov, Delay-independent stability and performance of distributed congestion control, IEEE/ACM Transaction on Networking, 15.
[19] K. Miller, T. Harks, Utility fair congestion control with time-varying delays, IEE INFOCOM, Phoenix, AZ, USA, 2008.
[20] S. Floyd, TCP and Explicit Congestion Notification, ACM Computer Communication Review 24 (5) (1994) 10–23.
[21] R.R.-F. Liao, P. Bouklee, A. Campbell, Online Generation of Bandwidth Utility Function for Digital Video, in: Proceedings of Packet Video'99, New York City, 1999.
[22] G. Bianchi, A. Campbell, R.R.-F. Liao, On utility-fair adaptive services in wireless networks, in: Proceedings of IEE/IFIP In. Workshop on QoS (IWQoS'98), 1998, Napa Valley, USA.
[23] T. Harks, T. Poschwatta, Priority pricing in utility fair networks, in: Proceedings of 13th IEEE International Conference on Network Protocols (ICNP), 2005, pp. 311–320.
[24] R. Gibbens, F. Kelly, Resource pricing and the evolution of congestion control, Automatica 35 (1999) 1969–1985.
[25] S. Low, A duality model of TCP flow controls, in: Proceedings of ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management, 2000.
[26] S. Athuraliya, V.H. Li, S.H. Low, Q. Yin, REM: Active Queue Management, IEEE Network 15 (2001) 48–53.
[27] S. Low, F. Paganini, J.C. Doyle, Scalable laws for stable network congestion control, in: Proceedings of Conference of Decision and Control, 2001, pp. 185–190.
[28] A. Tang, D. Wei, S.H. Low, M. Chiang, Heterogeneous congestion control: Efficiency, fairness and design, in: Proceedings of 14th IEEE International Conference on Network Protocols (ICNP), Santa Barbara, CA, 2006, pp. 127–136.
[29] R. Johari, J.N. Tsitsiklis, Efficiency loss in a network resource allocation game, Mathematical and Operational Research 29 (3) (2004) 407–435.
[30] T. Roughgarden, Selfish Routing and the Price of Anarchy, MIT Press, 2005.
[31] A. Tang, J. Wang, S. Low, Counter-intuitive throughput behaviors in networks under end-to-end control, IEEE/ACM Transactions on Networking 14 (2) (2006) 355–368.
[32] S.A. Vavasis, Complexity issues in global optimization: A survey, Tech. Rep., Ithaca, NY, USA, 1993.
[33] M. Bellare, P. Rogaway, The complexity of approximating a nonlinear program, Mathematical Programming 69 (1995) 429–442.
[34] S. McCanne, S. Floyd, The network simulator – ns-2, <http://www.isi.edu/nsnam/ns/>, 2000.

Tobias Harks received the Ph.D. degree in mathematics from University of Berlin, Germany, in 2007. He is currently working as a PostDoc at Technical University Berlin. His research interests include Combinatorial Optimization, Algorithmic Game Theory, Online Optimization, and Network Management.

Tobias Poschwatta received the Dipl.-Ing. degree in Electrical Engineering from Berlin University of Technology, Germany, in 2004. He is currently working as software engineer at Newtec Communications GmbH on protocols for Internet access via satellite.