# Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times

Milan Vojnović, Jean-Yves Le Boudec, and Catherine Boutremans
Institute for Computer Communications and Applications
Swiss Federal Institute of Technology at Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{Milan.Vojnovic, Jean-Yves.LeBoudec, Catherine.Boutremans}@epfl.ch

*Abstract*—**Consider a network with an arbitrary topology and arbitrary communication delays, in which congestion control is based on additive-increase and multiplicative-decrease. We show that the source rates tend to be distributed in order to maximize an objective function called $F_A^h$ ("$F_A^h$ fairness"). We derive this result under the assumption of rate proportional negative feedback and for the regime of rare negative feedback. This applies to TCP in moderately loaded networks, and to those TCP implementations that are designed to interpret multiple packet losses within one RTT as a single congestion indication and do not rely on re-transmission timeout. This result provides some insight into the distribution of rates, and hence of packet loss ratios, which can be expected in a given network with a number of competing TCP or TCP-friendly sources. We validate our findings by analyzing a multiple-bottleneck scenario, and comparing with previous results [1], [2], and an extensive numerical simulation with realistic parameter settings. We apply $F_A^h$ fairness to gain a more accurate understanding of the bias of TCP against long round-trip times.**

*Keywords*— **Additive-increase, Multiplicative-decrease, Fairness, Best-effort, TCP, TCP-friendly, TCP throughput-loss formula, RTT, Multiple-bottleneck, Stochastic approximation, ODE, Lyapunov.**

## I. INTRODUCTION

There is a continuing interest on throughput and fairness issues of TCP [3] congestion avoidance. This interest is particularly nourished by the proliferation of real-time "stream" applications over the Internet (e.g. voice, video) for which it is required to be TCP-friendly, i.e. to fairly coexist with already existing TCP applications.

In one of the pioneering works, Chiu and Jain [4] formulated a set of basic principles of the additive-increase and multiplicative-decrease congestion avoidance to achieve efficiency and fairness, by analyzing the simple model of a single bottleneck.

In [5] Kelly, Maulloo, and Tan showed that a large-scale network deploying some specific form of additive-increase and multiplicative-decrease congestion avoidance tends to distribute rates according to proportional fairness. This result is commonly misinterpreted as being applicable to congestion avoidance in the Internet with TCP.

Recently, Hurley, Le Boudec, and Thiran [6] showed that in a network employing additive-increase and multiplicative-decrease, the source rates tend to be distributed in order to maximize an objective function called $F_A$. The authors call this "$F_A$ fairness". This result is obtained by the limit mean ordinary differential equation (ODE) method, for a network operating in the regime of rare negative feedback. The pivotal assumption of that work is the rate proportional negative feedback, which the authors claim to be more realistic than one which depends exclusively on the overall load [5]. However, the result is restricted to the homogeneous round-trip time (RTT) case where the rates are updated synchronously.

In this paper, we extend the modeling of [6] to the heterogeneous RTT case. Our result is a generalization of $F_A$ fairness, which we call $F_A^h$ fairness. It gives the distribution of rates in an arbitrary network employing the additive-increase and multiplicative-decrease method for congestion control, with the assumption that negative feedback is rare. We allow the round-trip times to differ from one source to another. The rates tend to maximize an objective function called $F_A^h$, whose parameters reflect the rate adaptation algorithm. To the best of our knowledge, this is the first general result encompassing many of the relevant system parameters, applicable to an arbitrary network topology with multiple bottlenecks. Our results allows to find a first order approximation of rate distributions; combined with a loss-throughput formula such as [2], [7], this gives a prediction of the loss rates. The results are based on limit results in [8]; as such they apply only assimptotically to real network scenarios. However, our extensive simulations indicate that our results apply well in realistic cases.

The novelty of our approach is an application of the recent weak convergence results of decentralized asynchronous stochastic approximation algorithms [8]. Our model essentially differs from [6] in that we do not assume that rate-adaptation is performed synchronously by all sources; in contrast, we use an asynchronous model where every source updates its rate based on its own round-trip time interval. Unlike the synchronous model in [4], [5] or [6], this allows us to address the case with different round-trip times. But even in the case where all round-trip times are equal, this gives a more accurate model. Indeed, with the synchronous model, rate adjustment is based on the most recent previous rates. In reality, the feedback received by one source at the end of one round-trip time interval depends on the rates during the previous interval, shifted in time by the delay required for feedback to reach the sources. The synchronous model assumes implicitly that feedback reaches sources instantaneously. We call this assumption "stolen lag". We show with our modeling that the stolen lag assumption does not affect the distribution of average rates; by simulation, we see however that it affects the amplitude of oscillations. Note that our model explicitly considers all communication delays.

We assume in this paper that the negative feedback received by sources is rare, and is proportional to the source rate. The rare negative feedback assumption is valid in a reasonably loaded network; the proportional assumption should be true with active queue management [9] (e.g. RED [10]) applied to otherwise FIFO queues. In addition, our model assumes a single rate updating per RTT; this fits with TCP implementations designed to cope with multiple packet losses within single RTT, i.e. that treat multiple packet losses within one RTT as a single congestion signal, and avoid re-transmission timeouts. Also, it is assumed that all sources are greedy and long-lived.

Our model does not incorporate the effect of the variation of RTT for one given source from one feedback interval to the other. It is known that, for a network with fixed windows [11], the variation of round-trip times due to queues building up has in itself a congestion avoidance effect, which is not captured by our modeling. Another limitation is that we assume the rates to be piecewise constant, i.e. to be adjusted only once per round-trip time. Thus, the effect of burstiness at the timescale of the round-trip time is not taken into account. In contrast, our study captures the effect of the window or rate adaptation mechanism found for example with TCP or ABR. Our results may be used as a reference fairness measure in performance evaluations of TCP-friendly rate adjustment algorithms.

In the next subsection we outline our main results.

### A. Summary of the Main Results

We consider a network with multiple bottlenecks and heterogeneous round-trip times. Then, under the condition that there is no substantial queuing delay variation, and the network is operating in the regime of the rare negative feedback, the collection of rates $x = (x_1, ..., x_i, ...)$ is distributed such that $x$ maximizes the objective function

$$F_A^h(x) = \sum_{i \in \mathcal{S}} \frac{1}{\tau_i} \log \frac{x_i}{r_i + \eta_i x_i},$$

subject to the constraints $\sum_{j \in \mathcal{S}} A_{l,j} x_j \le c_l$, $\forall l \in \mathcal{L}$. In the formula, $\mathcal{S}$ is the set of sources, $\mathcal{L}$ the set of links, $A_{l,i}$ the routing matrix ($A_{l,i}$ is 0 or 1), $c_l$ the capacity of link $l$, and $\tau_i$ is the RTT for flow or source $i$. There is one flow per source. The rate adaptation parameters are $r_i$ (additive-increase element) and $\eta_i$ (multiplicative-decrease factor); they may depend on source $i$.

The above result is applied to a multiple-bottleneck network topology; we obtain a closed-form for the distribution of rates. This allows us to verify the consistency of our result with existing work and with conducted simulations. We find that the results in [1] are an asymptotic case of $F_A^h$ fairness for small additive-increase/multiplicative-decrease ratio relative to connection throughput.

We also gain a more accurate understanding of the bias of TCP against long round-trip times. We point out that it is important to make the difference between a bias against long RTTs (perhaps an undesirable feature) and a bias against flows with many hops (perhaps a desired feature). We see that the bias against flows with many hops is in the nature of any rate adaptation algorithm based on additive-increase and multiplicative-decrease. In contrast, a bias against long RTTs can be attenuated

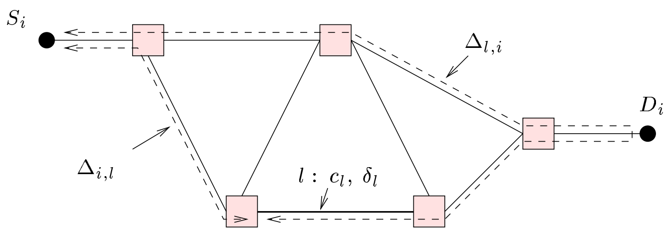



Fig. 1. An illustration of the defined delays.

with corrections such as mentioned in [1] and [12]. Finally, we also confirm throughput loss formulas, within the limitations of our modeling.

### B. Outline of the Paper

The paper is organized as follows. In Section II, the main results are derived. Following the basic model definitions, feedback modeling is described in more detail. Then, asymptotic convergence results of the decentralized asynchronous stochastic approximation algorithms [8] are sketched. In the rest of the section, objective function $F_A^h$ of the algorithm of concern is derived and analyzed. In Section III, $F_A^h$ result is applied to a multiple-bottleneck network topology for which a closed-form rate distribution is computed, and results are verified through numerical simulation. In Section IV, the results are discussed and compared to the related previous work. Implications of the result to the Internet are addressed in Section V. In Section VI, concluding remarks are given. In Appendix A we give the main theorem of the underlying theory [8].

## II. Derivation of the Main Results

### A. Model Setup

The notation is developed as follows. Let set $\mathcal{L}$ contain network links. Then let $c_l$ and $\delta_l$ be the capacity and delay of link $l \in \mathcal{L}$, respectively. Let set $\mathcal{S}$ comprise sources (flows) that are active on the given network. The routing setting we describe by routing matrix $A = (A_{l,i},\ l \in \mathcal{L},\ i \in \mathcal{S})$, such that $A_{l,i} = 1$, if flow $i$ traverses link $l$, and $A_{l,i} = 0$, otherwise.[1]

Further, we define communication delays. Let $\Delta_{i,l}$ denote delay from source $i$ to link $l$, and let $\Delta_{l,i}$ be the delay from link $l$ to source $i$. Then, set $\Delta_{i,j} = \Delta_{j,l} + \Delta_{l,i}$. Defining $\tau_i$ as the round-trip time of the source $i$, clearly, $\tau_i = \Delta_{i,l} + \Delta_{l,i}$, for all $l$ such that $A_{l,i} > 0$. In Fig. 1, a sample network illustrates defined delays.

Let $\{\tau_{i,n}\}_{n \ge 0}$ be a non-decreasing $[0, \infty)$-valued sequence of rate updating times of source $i$. Then, a number of rate updates of source $i$ on the interval $[0, t)$ is $N_i(t) = \sum_{n=1}^{\infty} 1_{\{\tau_{i,n} < t\}}$.

Let $\{x_{i,n}\}_{n \ge 0}$ be a $[0, \infty)$-valued stochastic process, where $x_{i,n}$ is a rate of source $i$ at the $n$-th update. Then, define a continuous time interpolation on real time as $x_i(t) = x_{i,n}$, for $t \in [\tau_{i,n}, \tau_{i,n+1})$.

For $b \ge a \ge 0$ define a $\sigma$-algebra of the form $\mathcal{F}^{i,l}_{[a,b)} = \sigma(x_{j,k} : A_{l,j}, A_{l,i} > 0, N_j(a - \Delta_{i,j}) \le k < N_j(b - \Delta_{i,j}))$. Finally, let $\mathcal{F}^{i}_{[a,b)} = \cup_{l:\ A_{l,i} > 0} \mathcal{F}^{i,l}_{[a,b)}$.

[1] We define $A_{l,i}$ on $\{0, 1\}$ which can be extended to $[0, 1]$ to accommodate for instance load sharing, etc. [6]

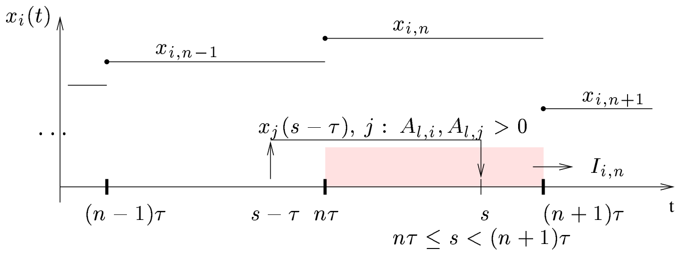



Fig. 2. Feedback modeling for the HOMRTT case.

An additive-increase and multiplicative-decrease algorithm has the following form

$$x_{i,n+1} = x_{i,n} + r_i(1 - I_{i,n}) - \eta_i I_{i,n} x_{i,n}, \tag{1}$$

where $r_i$ and $\eta_i$ are the additive-increase element and the multiplicative-decrease factor, respectively. The random sequence $\{I_{i,n}\}_{n\geq 0}$ is a negative feedback indication with values in $\{0, 1\}$. We assume that the negative feedback indication $I_{i,n}$ is based on the feedback received between $n$-th and $n+1$-th rate updating. Consequently, it turns out that $I_{i,n}$ is measurable on $\sigma$-algebra $\mathcal{F}^i_{[\tau_{i,n},\tau_{i,n+1})}$.

Now let us briefly comment on the special case addressed in [6], where it is assumed that all round-trip times are equal, and the rates are updated synchronously. Following the definition of $I_{i,n}$ in the full extent, one can see in Fig. 2 that $I_{i,n}$ is a function of $x_{j,n}$, $x_{j,n-1}$, and $x_{j,n-2}$, for all $j \in \mathcal{S}$ such that $A_{l,i}, A_{l,j} > 0$, depending on values of $\Delta_{j,l}$. In the related work [4]-[6], it is commonly assumed that $I_{i,n}$ is computed based on $x_{j,n}$, for all $j \in \mathcal{S}$ such that $A_{l,i}, A_{l,j} > 0$, which is indeed an unrealistic assumption.

Let us assume that all flows traversing link $l$ have equal access delay to that link. Formally, $\Delta_{i,l} = \Delta_{j,l}$, for all $i, j \in \mathcal{S}$, such that $A_{l,i}, A_{l,j} > 0$. Then, it follows that $I_{i,n}$ depends on $x_{j,n-1}$, for all $j \in \mathcal{S}$ such that $A_{l,i}, A_{l,j} > 0$. We refer to this assumption as a HOMRTT assumption. In addition, whenever $x_{i,n}$ is used where it should be $x_{i,n-1}$ we refer to this as a *stolen lag*. Feedback computation is illustrated in Fig. 2.

### B. Feedback Modeling

First, we introduce a notion of the link cost function $g_l(\cdot) : [0, \infty) \to [0, 1]$, where $l \in \mathcal{L}$. At a given time $t$, the link cost is a function of the link load

$$f_l(t) = \sum_{i\in\mathcal{S}} A_{l,i} x_i(t - \Delta_{i,l}). \tag{2}$$

One can interpret $g_l(f_l(t))$ as a probability of marking a single packet at time $t$. We are concerned with negative feedback indication, $I_{i,n}$, based on feedback received within $[\tau_{i,n}, \tau_{i,n+1})$. Let us partition the interval $[\tau_{i,n}, \tau_{i,n+1})$ into non-overlapping intervals $[a_k, b_k)$ such that $x_j(s)$, is constant for $s \in [a_k - \Delta_{i,j}, b_k - \Delta_{i,j})$, for all $i, j \in \mathcal{S}$ such that $A_{l,i}, A_{l,j} > 0$.

We define $M^{i,l}_{a,b}$ as an amount of negative feedback received by source $i$ from link $l$ within interval $[a, b)$, which is equal to a number of marked packets of the flow $i$ by the link $l$ within interval $[a - \tau_i, b - \tau_i)$.

Let $P^{i,l}_{a,b}(\cdot)$ and $P^i_{a,b}(\cdot)$ be conditional probabilities given $\mathcal{F}^{i,l}_{[a,b)}$ and $\mathcal{F}^i_{[a,b)}$, respectively. Analogously, let $E^{i,l}_{a,b}[\cdot]$ and $E^i_{a,b}[\cdot]$ be respective conditional expectations.

Admitting the interpretation of $g_l(\cdot)$ as a probability of marking a single packet, it is easy to see that we do have a binomial conditional probability

$$\begin{aligned} P^{i,l}_{a_k,b_k}(M^{i,l}_{a_k,b_k} = m) = \binom{\lceil x_{i,N_i(a_k-\tau_i)}(b_k - a_k)\rceil}{m} \cdot \\ \cdot g_l(f_l(a_k^{\Delta_i}))^m [1 - g_l(f_l(a_k^{\Delta_i}))]^{\lceil x_{i,N_i(a_k-\tau_i)}(b_k-a_k)\rceil - m}, \end{aligned} \tag{3}$$

where $a_k^{\Delta_i}$ is written in lieu of $a_k - \Delta_{l,i}$. Note that $\lceil x_{i,N_i(a_k-\tau_i)}(b_k - a_k)\rceil$ corresponds to a number of packets of flow $i$ that are present on link $l$ within $[a_k - \tau_i, b_k - \tau_i)$.

Clearly, the expected amount of negative feedback is

$$E^{i,l}_{a_k,b_k}[M^{i,l}_{a_k,b_k}] = \lceil x_{i,N_i(a_k-\tau_i)}(b_k - a_k)\rceil g_l(f_l(a_k^{\Delta_i})).$$

Similarly, let $M^i_{a,b}$ be an amount of negative feedback received within $[a, b)$ by source $i$ from all links $l$ such that $A_{l,i} > 0$. It follows

$$P^i_{a_k,b_k}(M^i_{a_k,b_k} = 0) = \prod_{l:\ A_{l,i}>0} P^{i,l}_{a_k,b_k}(M^{i,l}_{a_k,b_k} = 0),$$

and from (3) follows

$$\begin{aligned} &P^i_{a_k,b_k}(M^i_{a_k,b_k} = 0) = \\ &= \textstyle\prod_{l:\ A_{l,i}>0} [1 - g_l(f_l(a_k^{\Delta_i}))]^{\lceil x_{i,N_i(a_k-\tau_i)}(b_k-a_k)\rceil}. \end{aligned} \tag{4}$$

By definition of $[a_k, b_k)$ we have

$$P^i_{\tau_{i,n},\tau_{i,n+1}}(M^i_{\tau_{i,n},\tau_{i,n+1}} = 0) = \prod_k P^i_{a_k,b_k}(M^i_{a_k,b_k} = 0). \tag{5}$$

Finally, $I_{i,n} = 1$, if source $i$ has received an indication, within $[\tau_{i,n}, \tau_{i,n+1})$, that at least one packet has been marked, thus

$$\begin{aligned} &P^i_{\tau_{i,n},\tau_{i,n+1}}(I_{i,n} = 1) = P^i_{\tau_{i,n},\tau_{i,n+1}}(M^i_{\tau_{i,n},\tau_{i,n+1}} \geq 1) = \\ &= 1 - P^i_{\tau_{i,n},\tau_{i,n+1}}(M^i_{\tau_{i,n},\tau_{i,n+1}} = 0). \end{aligned} \tag{6}$$

Let us examine (6) for the HOMRTT case. Here we have a single partition of $[\tau_{i,n}, \tau_{i,n+1})$, hence, from (4)-(6) follows

$$P^i_{\tau_{i,n},\tau_{i,n+1}}(I_{i,n} = 1) = 1 - \prod_{l:\ A_{l,i}>0} [1 - g_l(f_l)]^{\lceil x_{i,n-1}\tau\rceil}, \tag{7}$$

where $\tau_{i,n+1} - \tau_{i,n} = \tau$, for all $i \in \mathcal{S}$, $n \geq 0$, and $f_l$ stands for $f_l(\tau_{i,n} - \Delta_{l,i}) = \sum_{j\in\mathcal{S}} A_{l,j} x_{j,n-1}$. In the limit case $g_l(\cdot) \to 0$, limited development yields $[1 - g_l(f_l)]^{\lceil x_{i,n-1}\tau\rceil} \simeq 1 - \lceil x_{i,n-1}\tau\rceil g_l(f_l)$, then replacing this in (7), and neglecting the higher order products, yield

$$P^i_{\tau_{i,n},\tau_{i,n+1}}(I_{i,n} = 1) \simeq \sum_{l\in\mathcal{L}} A_{l,i} g_l(f_l) \lceil x_{i,n-1}\tau\rceil. \tag{8}$$

Therefore, it is shown that, under the rare negative feedback assumption, (7) degenerates to the rate proportional feedback as is implicitly assumed in [6]. However, note that (8) depends on $x_{i,n-1}$ and not on $x_{i,n}$.

## C. Asymptotic Convergence

Traditional theory of the stochastic approximation algorithms [13]-[14] is concerned with an algorithm of the general form

$$x_{i,n+1} = x_{i,n} + \gamma_n H_{i,n}(x_{i,n}, \xi_{i,n}),\ i \in \mathcal{S},$$

where $x_{i,n}$ is defined on $R$, $H_{i,n}(\cdot): R \times R \to R$, $\xi_{i,n}: R \to R$ is a random noise, and $\gamma_n$ a step size. It is assumed that components $x_{i,\cdot}$, $i \in \mathcal{S}$, are updated synchronously, facilitating the association of continuous interpolation $x_i(t)$ to discrete process $\{x_{i,n}\}_{n\geq 0}$ on the "natural" common iterate time $\{n\gamma\}_{n\geq 0}$. However, it follows that for an asynchronous updating one has to work in real time, or at least an appropriately scaled real time [8]. In general, for decreasing $\gamma_n$, in respect to $n$, convergence with probability one can be obtained, while for constant small $\gamma_n \equiv \gamma$, only convergence in probability can be proven (weak convergence). The weak convergence limit is obtained for a small enough step size $\gamma$, and observing interpolation of discrete time process of concern on the real time scaled by $\gamma$. In the rest of this section we briefly sketch results of [8] that are applied in our work. For a complete treatment of the underlying theory the reader is referred to [8].

Let $\{\delta\tau^\gamma_{i,k}\}_{k\geq 0}$ be a random sequence of updating intervals of $\{x_{i,k}\}_{k\geq 0}$, $i \in \mathcal{S}$. Then denote the (scaled) real updating time of $x_{i,n}$ as

$$\tau^\gamma_{i,n} = \gamma \sum_{k=0}^{n-1} \delta\tau^\gamma_{i,k}, \tag{9}$$

and a continuous interpolation on the iterate time $\tau^\gamma_i(t) = \tau^\gamma_{i,n}$, for $t \in [n\gamma, (n+1)\gamma)$. Further, let $N^\gamma_i(t)$ is a number of updates of $\{x_{i,k}\}_{k\geq 0}$ before $t/\gamma$. Formally,

$$N^\gamma_i(t) = \gamma \sum_{n=1}^{\infty} 1_{\{\tau^\gamma_{i,n} < t/\gamma\}}. \tag{10}$$

From the definitions, it turns out that $N^\gamma_i(\tau^\gamma_i(t)) = n\gamma$, $t \in [n\gamma, (n+1)\gamma)$, i.e. $N^\gamma_i(\cdot)$ is inverse of $\tau^\gamma_i(\cdot)$.

Then, let $\hat{x}^\gamma_i(t) = x^\gamma_{i,n}$, $t \in [\tau^\gamma_{i,n}, \tau^\gamma_{i,n+1})$, is a continuous interpolation on scaled real time, and $\hat{x}^\gamma(\cdot) = (\hat{x}^\gamma_i(\cdot), i = 1, 2, \ldots, S)$, $S = |\mathcal{S}|$. From definitions of "time" processes (9) and (10) it follows $x^\gamma_i(t) = \hat{x}^\gamma_i(\tau^\gamma_i(t))$, and $\hat{x}^\gamma_i = x^\gamma_i(N^\gamma_i(t))$.

Furthermore, let $\Delta^\gamma_{i,j,n}$ be a non-negative random variable representing a scaled (multiplied by $\gamma$) communication delay between sources $i$ and $j$ at the $n$-th rate updating of source $i$.

Finally, the decentralized asynchronous algorithm can be written in the form

$$\begin{aligned} x^\gamma_{i,n+1} &= \Pi_{[a_i,b_i]}\left(x^\gamma_{i,n} + \gamma H^\gamma_{i,n}(\hat{x}_j(\tau^\gamma_{i,n+1} - \Delta^\gamma_{i,j,n}))\right) = \\ &= x^\gamma_{i,n} + \gamma H^\gamma_{i,n}(\hat{x}_j(\tau^\gamma_{i,n+1} - \Delta^\gamma_{i,j,n})) + \gamma Z^\gamma_{i,n},\ i \in \mathcal{S}, \end{aligned} \tag{11}$$

where $\Pi_{[a_i,b_i]}(\cdot)$ denotes projection of the argument on $[a_i, b_i]$, for the constrained $x$ on $C = [a_1,b_1] \times [a_2,b_2] \times \ldots \times [a_S,b_S]$, and $Z^\gamma_{i,n}$ is a reflection term.

Let for all $i$, $\mathcal{F}^\gamma_{i,n}$ and $\mathcal{F}^{\gamma,+}_{i,n}$ be non-decreasing $\sigma$-algebras measuring the past data (including $x^\gamma_{i,0}$, $H^\gamma_{j,k}$, and $\delta\tau^\gamma_{j,k}$, $j \in \mathcal{S}$) available on $[0, \tau^\gamma_{i,n+1})$, and $[0, \tau^\gamma_{i,n+1}]$, respectively. Then, with $P^\gamma_{i,n}$ and $P^{\gamma,+}_{i,n}$ denote respective conditional probabilities, and analogously $E^\gamma_{i,n}$ and $E^{\gamma,+}_{i,n}$ conditional expectations.

Subsequently, we have the following conditions. It is assumed that

$$\{H^\gamma_{i,n}, \delta\tau^\gamma_{i,n}; \gamma, i, n\}, \tag{12}$$

and $\Delta^\gamma_{i,j,n}$, $\Delta^{\gamma,+}_{i,j,n}$ are uniformly integrable. We consider the martingale difference noise [8], for which we have

$$E^\gamma_{i,n} H^\gamma_{i,n} = h^\gamma_{i,n}(\hat{x}^\gamma_j(\tau^\gamma_{i,n+1} - \Delta^\gamma_{i,j,n}), j \in \mathcal{S}) + \beta^\gamma_{i,n},$$

where $h^\gamma_{i,n}(\cdot)$ are real-valued functions continuous in $n$ and $\gamma$, $\beta^\gamma_{i,n}$ is asymptotically negligible noise, and $\sup_{n\leq T/\gamma} \Delta^\gamma_{i,j,n} \to 0$. There are real-valued functions $u^\gamma_{i,n}(\cdot)$ that are strictly positive ($\inf_{n,\gamma,x,\xi} u^\gamma_{i,n}(x,\xi) > 0$) and are continuous uniformly in $n$ and $\gamma$, and non-negative random variables $\Delta^{\gamma,+}_{i,j,n}$ such that

$$E^{\gamma,+}_{i,n} \delta\tau^\gamma_{i,n+1} = u^\gamma_{i,n+1}(\hat{x}^\gamma_j(\tau^\gamma_{i,n+1} - \Delta^{\gamma,+}_{i,j,n+1}), j \in \mathcal{S}), \tag{13}$$

where $\sup_{n\leq T/\gamma} \Delta^{\gamma,+}_{i,j,n} \to 0$, in probability as $\gamma \to 0$. There are continuous real-valued functions $\bar{h}_i(\cdot)$ such that for each $x \in C$,[2]

$$\lim_{m,n,\gamma} \frac{1}{m} \sum_{k=n}^{n+m-1} \gamma[h^\gamma_{i,k}(x) - \bar{h}_i(x)] = 0. \tag{14}$$

There are continuous real-valued functions $\bar{u}_i(\cdot)$ such that for all $x \in C$,

$$\lim_{m,n,\gamma} \frac{1}{m} \sum_{k=n}^{n+m-1} E^{\gamma,+}_{i,n}[u^\gamma_{i,k}(x) - \bar{u}_i(x)] = 0. \tag{15}$$

Suppose

$$\lim_{m,n,\gamma} \sum_{k=n}^{n+m-1} E^\gamma_{i,n} \beta^\gamma_{i,k} = 0, \text{ in mean.} \tag{16}$$

Finally, from the Theorem [8] (Appendix A) particularly follows that, for the unconstrained algorithm, the weak convergence subsequence $\hat{x}$ is the limit set of ODE

$$\dot{\hat{x}}_i = \frac{\bar{h}_i(\hat{x})}{\bar{u}_i(\hat{x})},\ i \in \mathcal{S}. \tag{17}$$

Thus, the limit mean ODE is the same as in the synchronous case, except for an additional weight factor that takes into account frequency of the updating.

## D. $F^h_A$ Fairness

We identify $H^\gamma_{i,n}$ of (11) in the algorithm (1) as

$$H^\gamma_{i,n} = r_i - (r_i + \eta_i x_{i,n}) 1_{\{\xi_{i,n} < P^i_{i,n+1}(I^\gamma_{i,n}=1)\}}, \tag{18}$$

where by definition $\{\xi_{i,n}\}_{n\geq 0}$ is a sequence of independent random variables uniformly distributed on $[0,1]$.[3] Then,

$$E^\gamma_{i,n+1} H^\gamma_{i,n} = r_i - (r_i + \eta_i x_{i,n}) P^\gamma_{i,n+1}(I^\gamma_{i,n} = 1),$$

[2] In (14), (15), and (16), $\lim_{m,n,\gamma} \equiv \lim_{m\to\infty, n\to\infty, \gamma\to 0}$, simultaneously in any way.

[3] Note that instead of $r_i$ and $\eta_i$ it should be written $r^\gamma_i$ and $\eta^\gamma_i$, where $r_i = \gamma r^\gamma_i$ and $\eta_i = \gamma\eta^\gamma_i$, but we abuse this for notation simplicity.

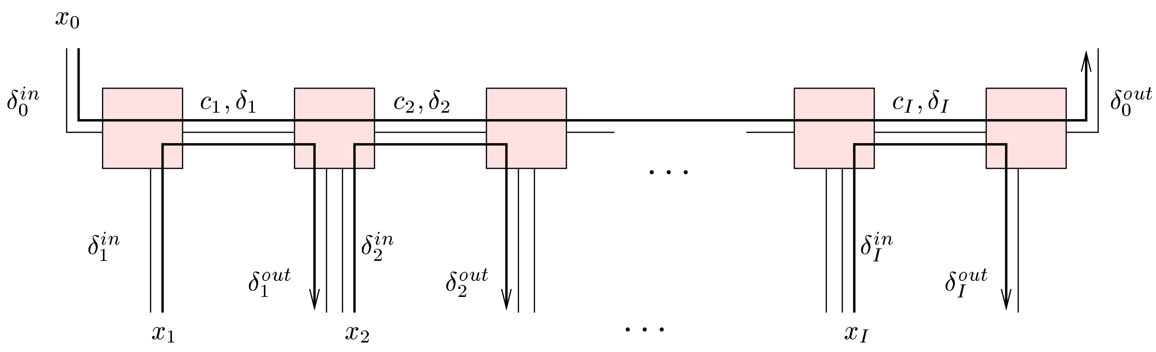



Fig. 3. A multiple-bottleneck network topology.

where $P^\gamma_{i,n+1}(I^\gamma_{i,n} = 1)$ is given by (6).

In the limit case, as $\gamma \to 0$, and $n \to \infty$, we neglect scaled delays $\Delta_{i,j}$, then the probability of negative feedback is

$$P^\gamma_{i,n+1}(I^\gamma_{i,n} = 1) = \hat{x}_i(\tau^{\gamma,-}_{i,n+1})\delta\tau_{i,n} \sum_{l\in\mathcal{L}} A_{l,i} g_l(\hat{f}_l(\tau^{\gamma,-}_{i,n+1})).$$

For $\delta\tau_{i,n} = \tau_i$, for all $n > 0$, the mean vector field is

$$\bar{h}_i(\hat{x}(t)) = r_i - \hat{x}_i(t)(r_i + \eta_i \hat{x}_i(t))\tau_i \sum_{l\in\mathcal{L}} A_{l,i} g_l(\hat{f}_l(t)) \quad (19)$$

where $\hat{f}_l(\cdot) = \sum_{j\in\mathcal{S}} A_{l,j}\hat{x}_j(\cdot)$.

In our case the rate is updated at the end of each round-trip time interval, thus, $\bar{u}_i(\hat{x}) = \tau_i$. Then, combining this with (19) the limit mean ODE (17) becomes

$$\dot{\hat{x}}_i = \frac{r_i}{\tau_i} - \hat{x}_i(r_i + \eta_i\hat{x}_i) \sum_{l\in\mathcal{L}} A_{l,i} g_l(\hat{f}_l). \quad (20)$$

Following the same steps as in [6] we express ODE (20) as

$$\dot{\hat{x}}_i = \hat{x}_i(r_i + \eta_i\hat{x}_i) \frac{\partial J^h_A(\hat{x})}{\partial \hat{x}_i}, \quad (21)$$

where

$$J^h_A(\hat{x}) = \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log \frac{\hat{x}_i}{r_i + \eta_i\hat{x}_i} - G(\hat{x}), \quad (22)$$

and by definition $G(\hat{x}) = \sum_{l\in\mathcal{L}} G_l(\hat{f}_l)$, where $G_l(\cdot)$ is a primitive of $g_l(\cdot)$.

It is easy to see that $J^h_A$ is strictly concave and consequently has a unique maximum over any bounded region. It turns out that $J^h_A$ is Lyapunov for ODE and with an unique attractor, for which $J^h_A$ is maximized.

Along the same lines as in [6], one can neglect the second term in (22). Then, it follows that the rates $x$ are distributed such that $x$ maximizes

$$F^h_A(x) = \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log \frac{x_i}{r_i + \eta_i x_i} \quad (23)$$

subject to the constraints

$$\sum_{j\in\mathcal{S}} A_{l,j} x_j \le c_l, \ \forall l \in \mathcal{L}. \quad (24)$$

D.1 $F^h_A$ asymptotic limits

Let us write (23) in the following form

$$F^h_A(x) = \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log \frac{1}{\eta_i} - \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log\left(1 + \frac{r_i}{x_i\eta_i}\right). \quad (25)$$

Then we develop the second term to obtain

$$F^h_A(x) = \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log \frac{1}{\eta_i} - \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \sum_{n=1}^{\infty} \frac{(-1)^{\mathrm{n}-1}}{n} \left(\frac{r_i}{\eta_i x_i}\right)^n,$$

for $x_i > r_i/\eta_i$. By the limited development for $x_i >> r_i/\eta_i$ we obtain that rates $x$ are distributed such that $x$ maximizes

$$F^{h+}_A(x) = -\sum_{i\in\mathcal{S}} \frac{r_i}{\tau_i\eta_i x_i}, \quad (26)$$

subject to (24).[4] Note that we skip the first term in (25) that is not relevant for maximization in respect to $x$. In the opposite case, $x_i << r_i/\eta_i$, for all $i \in \mathcal{S}$, by simple manipulation we obtain an objective function

$$F^{h-}_A(x) = \sum_{i\in\mathcal{S}} \frac{1}{\tau_i} \log x_i. \quad (27)$$

We use (26) and (27) in Section V to understand the bias of TCP against long RTTs.

## III. An Example of $F^h_A$ Fairness – multiple-bottleneck

In this section we consider a multiple-bottleneck topology depicted in Fig. 3. In this scenario we distinguish class 0 flows, traversing tandem of $I$ links, and class $i$ flows traversing a single link $i$, $i = 1, 2, \ldots, I$. Let $n_i$ be the number of flows belonging to class $i$.

Feasibility constraints $n_0x_0 + n_ix_i = c_i$, $i = 1, 2, \ldots, I$, allow us to express (23) in terms of $x_0$ as follows

$$F^h_A(x_0) = \frac{n_0}{\tau_0} \log \frac{x_0}{r_0 + \eta_0 x_0} + \sum_{i=1}^{I} \frac{n_i}{\tau_i} \log \frac{c_i - n_0x_0}{r_in_i + \eta_i(c_i - n_0x_0)}. \quad (28)$$

[4]For TCP congestion avoidance $r_i = 1/\tau_i$ and $r_i = 1/(2\tau_i)$ for not delayed ACKs and delayed ACKs, respectively. Henceforth, (26) can be safely applied to the respective cases as long as the number of packets per RTT, $x_i\tau_i >> 2$, and $x_i\tau_i >> 1$, for all $i \in \mathcal{S}$.

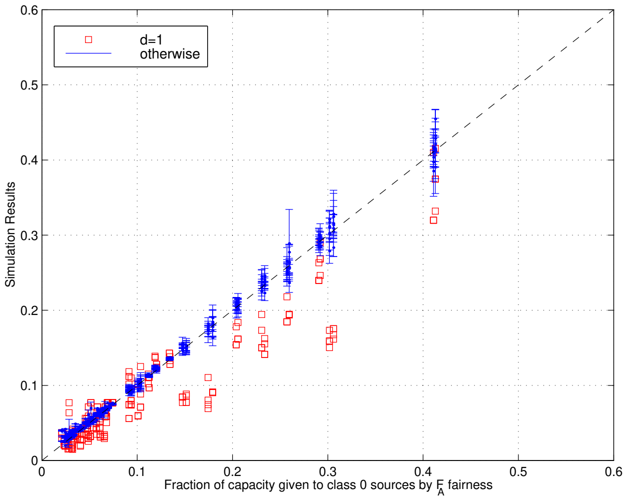



Fig. 4. HOMRTT with the stolen lag, $\tau_i = 0.2$ s, $r_i = 5$, $\eta_i = 0.5$ for all flows.

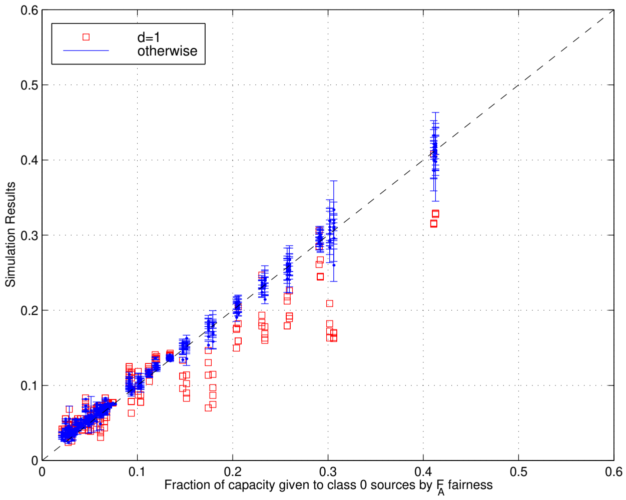



Fig. 5. HOMRTT with no stolen lag, $\tau_i = 0.2$ s, $r_i = 5$, $\eta_i = 0.5$ for all flows.

In the sequel, we suppose

$$c_i = c, r_i = r_I, \eta_i = \eta_I, n_i = w, i = 1, 2, \ldots, I, n_0 = v. \tag{29}$$

Then, (28) becomes

$$F_A^h(x_0) = \frac{v}{\tau_0} \log \frac{x_0}{r_0 + \eta_0 x_0} + T_I w \log \frac{c - v x_0}{r_I n_I + \eta_I (c - v x_0)}, \tag{30}$$

where $T_I \triangleq \sum_{i=1}^{I} \frac{1}{\tau_i}$.

*Lemma 1* (multiple-bottleneck) $F_A^h$–fairness distribution for the multiple-bottleneck scenario with (29) is

$$x_0 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}, \tag{31}$$

where

$$\begin{aligned} A &= v^2 r_0 \eta_I - w^2 T_I r_I \eta_0 \tau_0 \\ B &= -r_0 (v(w r_I + 2c\eta_I) + w^2 T_I r_I \tau_0) \\ C &= r_0 c (w r_I + c \eta_I), \end{aligned}$$

for $v^2 r_0 \eta_I - w^2 r_I \eta_0 T_I \tau_0 \neq 0$, otherwise

$$x_0 = \frac{c(w r_I + c\eta_I)}{v(w r_I + 2c\eta_I) + w^2 T_I r_I \tau_0}. \tag{32}$$

Then,

$$x_i = \frac{c - v x_0}{w},$$

for $i = 1, 2, \ldots, I$.

*Proof:* Proof is simple and is same as in [6]. ■

Subsequently, with $x_0 >> r_0/\eta_0$ and $x_i >> r_I/\eta_I$, for each $i = 1, 2, \ldots, I$, the first-order approximation (26) yields[5]

$$\frac{x_0}{c} = \frac{1}{v + w\sqrt{T_I \frac{r_I \eta_0 \tau_0}{r_0 \eta_I}}}, \tag{33}$$

or in a simplified case $\tau_i = \tau_I$, for $i = 1, 2, \ldots, I$, so that $T_I = \frac{I}{\tau_I}$,

$$\frac{x_0}{c} = \frac{1}{v + w\sqrt{I \frac{r_I \eta_0 \tau_0}{r_0 \eta_I \tau_I}}}. \tag{34}$$

TABLE I

FRACTION OF CAPACITY $c$ GIVEN TO A CLASS 0 FLOWS, FOR THE MULTIPLE-BOTTLENECK, WITH $x_i >> r_i/\eta_i$ FOR ALL $i \in \mathcal{S}$.

| Fairness | $x_0/c$ | parameters setup |
|---|---|---|
| $F_A$ | $\frac{1}{v+w\sqrt{I}}$ | $r_0 = r_I, \tau_0 = \tau_I; r_0 = K\tau_0, r_I = K\tau_I$ |
| Proportional | $\frac{1}{v+wI}$ | $r_0 = r_I, \tau_0 = I\tau_I$ |
| Max-min | $\frac{1}{v+w}$ | $r_0 = KI\tau_0, r_I = K\tau_I$ |
| TCP-Reno | $\frac{1}{v+wI\sqrt{I}}$ | $r_0 = 1/\tau_0, r_I = 1/\tau_I, \tau_0 = I\tau_I$ |

*Remarks*: Row 1 – $F_A$-like result [6] is achieved for homogeneous RTTs, and additive-increase parameters proportional to the RTTs. Row 2 – Proportional fairness result holds for the HETRTT network and equal values of $r_i$ and $\eta_i$ for all sources. Row 3 – Max-min fairness can be obtained by indicated parameters setting. Row 4 – Indicates rate distribution for TCP Reno and the HETRTT network (def. Section III-A), which complies to the other results mentioned in Section III-A.

Finally, we consider a few special cases for some reasonable parameters setup as is given in Table I. It can be observed that various fairness objectives are achieved depending on the end-system parameters and RTTs.

## *A. Simulation Results*

We perform an extensive numerical simulation study of the multiple-bottleneck scenario shown in Fig. 3. Let $\delta_i^{in}$ and $\delta_i^{out}$ be access delays of class $i$ flow, as depicted in Fig. 3. Note that the definition of the multiple-bottleneck encompasses both a single bottleneck with different RTTs ($I = 1$), and a multiple bottleneck case, as is considered separately in [1]. To facilitate computation, all delays and rate updating intervals are set as multiples of $\varepsilon > 0$. All events corresponding to the rate updating are aligned in respect to the sequence of $\varepsilon$ slots. We use the same form of the link cost function as is suggested in [6]

$$g_l(f_l) = \begin{cases} 0, & f_l < d \\ \left(\frac{f_l/c - d}{1-d}\right)^p, & d \le f_l \le 1 \\ 1, & f_l > 1 \end{cases},$$

[5] In [6] it is referred to this case as $\lim_{c\to\infty} x_0/c$, which is encompassed in $x_i >> r_i/\eta_i$, for $r_i < \infty$ and $\eta_i > 0$.

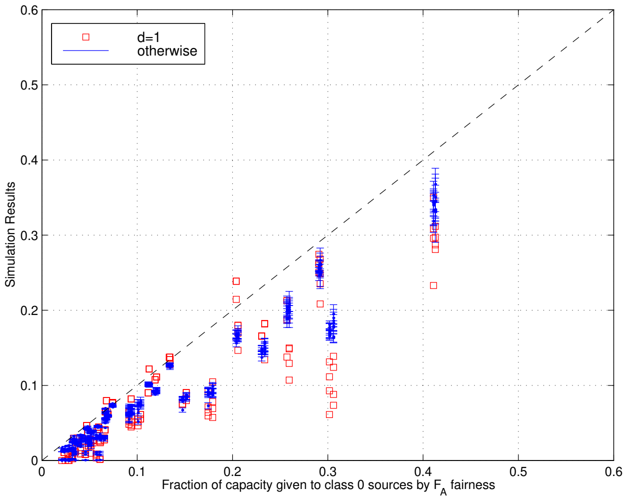



Fig. 6. HETRTT, $F_A$ vs. simulation, $\tau_0 = 0.2$ s, $\tau_i = \tau_0/I$, for $i = 1, 2, \ldots, I$, and $r_i = 5$, $\eta_i = 0.5$ for all flows.

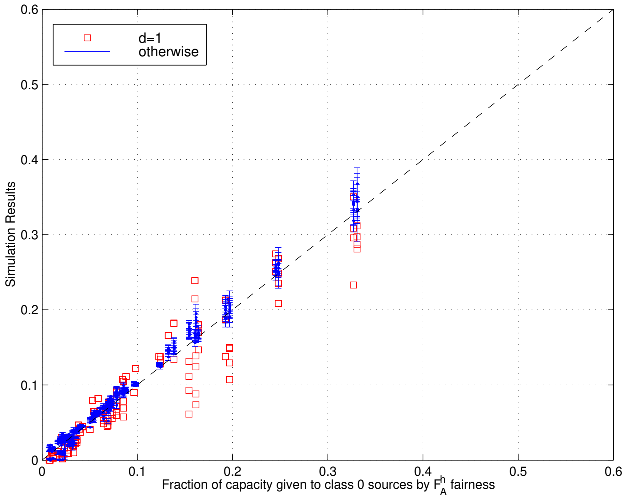



Fig. 7. HETRTT, $F_A^h$ vs. simulation, $\tau_0 = 0.2$ s, $\tau_i = \tau_0/I$, for $i = 1, 2, \ldots, I$, and $r_i = 5$, $\eta_i = 0.5$ for all flows.

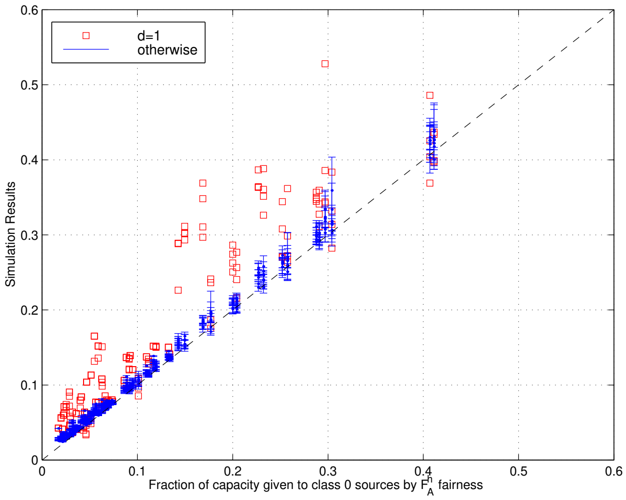



Fig. 8. HETRTT, additive-increment proportional to RTT, $r_i = K\tau_i$, $K = 25$, $\eta_i = 0.5$, $\tau_0 = 0.2$ s, $\tau_i = \tau_0/I$, $i = 1, 2, \ldots, I$.

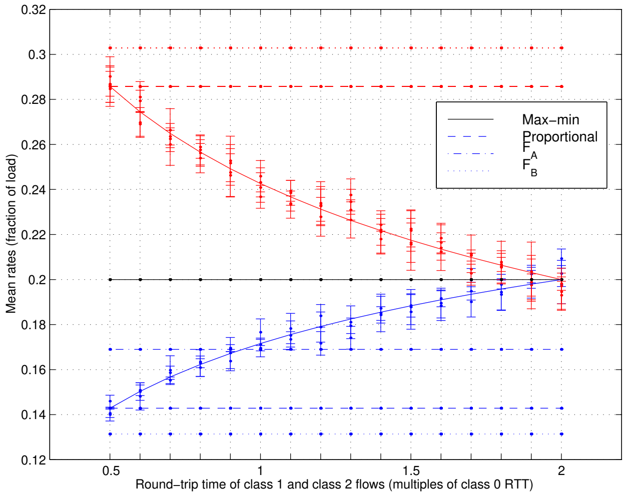



Fig. 9. Impact of the RTT on rate distribution, $c = 250$, $I = 2$, $v = 3$, $w = 2$, $\tau_0 = 0.2$, $\tau_i = 0.1 - 0.4$, $\delta_1 = \delta_2 = \delta = 5\varepsilon = 0.05$.

where $c$ is a link capacity, $d \in [0, 1]$, and $p > 0$.

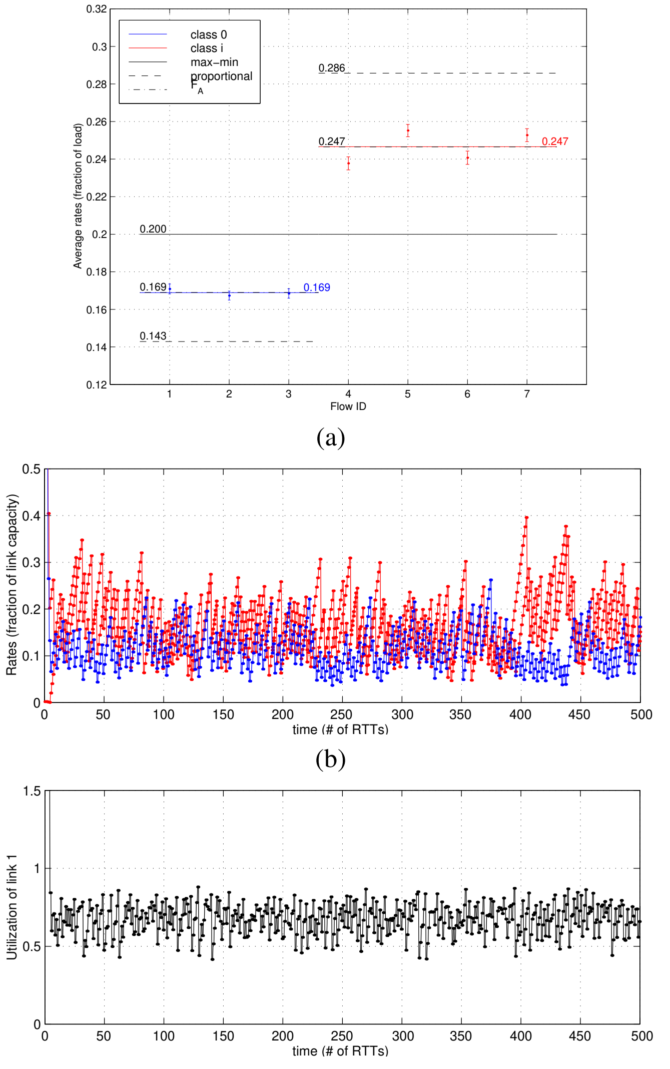
We consider all combinations of the simulation parameters that are set as follows: $I = \{2, 5\}$, $v, w = \{1, 2, 6, 12\}$, $c = \{250, 625\}$, $d = \{0, 0.5, 1\}$, and $p = \{1, 2, 5, 10\}$. Hence, there are a total of 768 settings, and we believe that this covers a wide range of realistic situations. In addition, it is noteworthy that all simulations are obtained for TCP-like parameters, and not for the limit small values. Total simulation time is set equal to 500 the largest RTT. Each average value is obtained over four simulation runs, excluding the initial 20% of the trace to eliminate initial transient. All confidence intervals are computed as 95% of confidence. For all simulation results, the claim of [6] that the results substantially deviate for the $d = 1$ case is confirmed, therefore, these results are plotted differently. Hereinafter, we refer to the multiple-bottleneck with all access delays equal to zero as a HETRTT; formally $\delta_i^{in} = \delta_i^{out} = 0$, for all $i = 0, 1, \ldots, I$. In the sequel, scatter plots of the $F_A^h$ analytical throughput against the corresponding simulation result for class 0 flows are shown.

In Fig. 4 and 5 the HOMRTT network is considered with stolen lag and no stolen lag assumption, respectively. The simulations confirm what follows from the model (Section II-D) that rates are distributed according to $F_A^h$ regardless wheather the feedback is based on the last rate updating or on the value that is one lag behind.

In Fig. 6 and 7, for the HETRTT scenario, the $F_A$ and $F_A^h$ analytical results are plotted against respective simulation counterparts. It is evident that the $F_A^h$ matches the simulation results by taking into account diversity of the RTTs.

Fig. 8 demonstrates how the additive-increase proportional to the RTT [1], for the HETRTT network, corrects the bias against the connections with long RTTs, thus the simulation results conform to both $F_A$ and $F_A^h$ fairness.

In Fig. 9 we consider the multiple-bottleneck network with 2 links, such that all access delays are zero except the egress link of class 1 flows and ingress link of class 2 flows, which are set

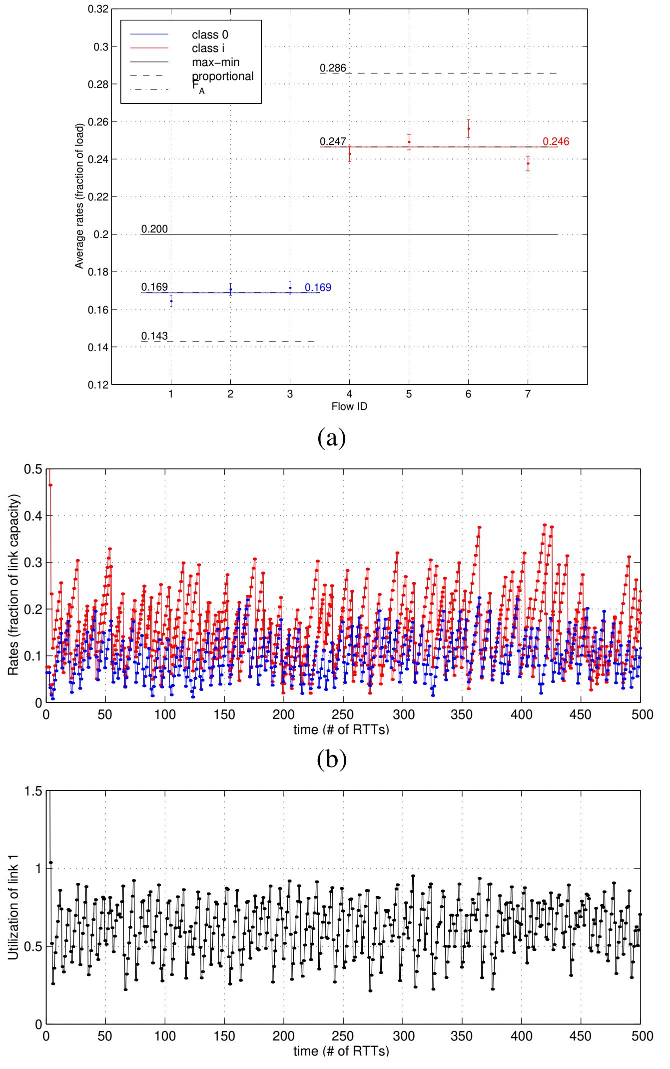



(a)

(b)

(c)

Fig. 10. HOMRTT with the stolen lag, $c = 250$, $\tau = 0.2$, $r_i = 5$, $\eta_i = 0.5$, $I = 2$, $v = 3$, $w = 2$, $d = 0.5$, $p = 5$, (a) rate distribution, (b) rate traces, and (c) utilization of link 1.




(a)

(b)

(c)

Fig. 11. HOMRTT with no stolen lag, $c = 2500$, $\tau = 0.2$, $r_i = 5$, $\eta_i = 0.5$, $I = 2$, $v = 3$, $w = 2$, $d = 0.5$, $p = 5$, (a) rate distribution, (b) rate traces, and (c) utilization of link 1.

equal and varied such that the network setting gradually shifts from the HETRTT case (the leftmost point, Fig. 9) to the case where $\tau_1$ and $\tau_2$ are double of $\tau_0$ (the rightmost point, Fig. 9). The results substantiate the validity of $F_A^h$ and illustrate how the system operates according to different fairness criteria depending on RTT.

## IV. Discussion and Related Work

To simplify the presentation let us consider the HOMRTT network with synchronous rate updating and common RTT, $\tau$, where $\{x_k\}_{k\geq 0}$ represents a sequence of rate vectors. In Section II-A we argued that $n+1$-th rate updating $x_{n+1}$ is based on the feedback depending on $x_{n-1}$, and not on $x_n$, as is commonly assumed in the existing work, e.g. [4]-[6]. Recall that we refer to this as the stolen lag. From the weak convergence it follows that the stolen lag is not relevant in respect to the limit mean rate distribution. However, the rate dynamics with no stolen lag has a higher variance.

Intuitively, this can be explained as follows. Let us assume that the network operates in the steady state and the packet-loss ratio is small and constant. Then, negative feedback $I_{i,n}$ is mainly driven by either $x_{i,n}$ or $x_{i,n-1}$, for the stolen lag and no stolen lag, respectively. During the period of rate decreasing, at $n+1$-th rate updating, rate $x_{i,n}$ is further decreased with probability proportional to either $x_{i,n}$, or $x_{i,n-1}$, for the stolen lag and no stolen lag, respectively. Since $x_{i,n-1} \geq x_{i,n}$, it follows that with a higher probability the rate is further decreased with no stolen lag, compared to the stolen lag case. Analogous reasoning applies also for a period of the rate increasing. This implies a higher variance of the rate for the case with no stolen lag. We substantiated the above argument by Markov chain modeling of a discrete process $x$, for which we computed stationary distribution, for both cases. However, due to the space limitations these results are not shown here. Instead, in Fig. 10 and 11, the HOMRTT simulation results are shown for both the stolen lag and no stolen lag case, respectively. Traces of the utilization of

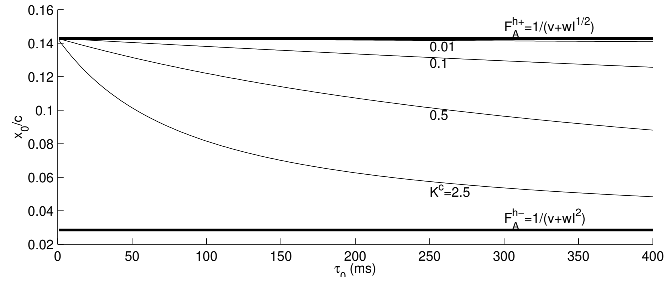



Fig. 12. Residual bias against long RTT connections, HETRTT, $\tau_I = \tau_0/I$, $I = 4, v = 3, w = 2, K^c = K/c$.

link 1, in Fig. 10c and 11c, particularly demonstrate observed phenomena.

Furthermore, it turns out that for the multiple-bottleneck topology, analyzed in Section III, the throughput of flows traversing multiple bottlenecks (34) fully conforms to the results obtained in [1]. Let $W(\cdot)$ be a real-valued strictly positive function, then for TCP it generally holds $r_i = W(\tau_i)/\tau_i$, thus (34) becomes

$$\frac{x_0}{c} = \frac{1}{v + w\sqrt{I}\frac{\tau_0}{\tau_I}\sqrt{\frac{W(\tau_I)}{W(\tau_0)}}},$$

which is exactly the same form, for both a single bottleneck ($I = 1$) with different RTTs, and multiple bottlenecks, as obtained by simplified modeling in [1].

## V. Application to the Internet

### A. Bias Against Connections with Long RTTs

Our fundamental result allows us to better understand the bias of TCP against connections with long round-trip times.

First, for any congestion control mechanism, if the distribution of rates tends to maximize a concave utility function, then flows with many hops are likely to receive a small rate [5]. Since $F_A^h$ is concave with $x_i$, this is true with our system, whatever the rate adaptation parameters $r_i$ and $\eta_i$ are. This is probably a desired bias, since flows with many hops use more network resources. In practice, many hops often mean larger RTT, but not always.

Second, both the specific values of the rate adaptation parameters, $r_i$ and $\eta_i$, and the update frequency $1/\tau_i$ also play a role ($\tau_i$ is the RTT for source $i$). With TCP-Reno, with no delayed ACKs, we have:

- $r_i = \mathrm{MSS}_i/\tau_i$ – in the congestion avoidance phase, with no delayed ACKs, the window is effectively increased by one packet per RTT (resp. $1/2$ packet for delayed ACKs). In the formula, $\mathrm{MSS}_i$ is the packet size for source $i$.
- $\eta_i = 0.5$ – when a loss is detected, the target window size is divided by a factor 2.

This results in an obvious bias against long RTTs: the increase element $r_i$ is smaller, and less frequent for long RTTs.

A fix to this undesired bias has been proposed in [1]; it consists in setting the increase element to $r_i = K\tau_i$, where $K$ is a common constant. This is actually done by multiplying the window increase parameter by $\tau_i^2$, considering that the rate is equal

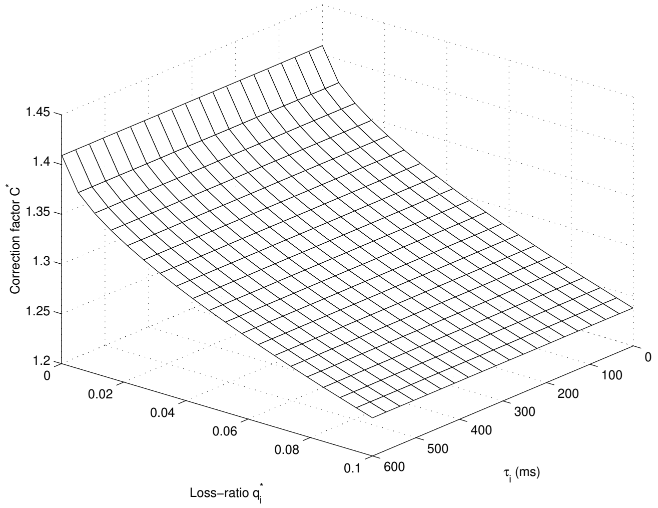



Fig. 13. Correction of the TCP throughput-loss formula (not delayed ACKs).

to the window size divided by $\tau_i$. In the limiting case where rates $x_i$ are large, we can apply the first order development (26) and obtain that the rates tend to maximize $-\sum_{i\in\mathcal{S}}\frac{1}{x_i}$, thus there is no bias against long RTTs in this case (but the bias against many hops does exist). However, the non-linear nature of $F_A^h$ indicates that this is not always true. If some rates are small, for example in a very loaded network, then their utility function equals to $\frac{1}{\tau_i}(-\log(K\tau_i) + \log x_i)$, or if this holds for all flows, then the distribution of rates tends towards weighted proportional fairness (27), with a weight equal to the inverse of the RTT. Thus there might still remain an undesirable negative bias against long RTTs for precisely those sources that are not able to obtain a large rate. We can verify this bias in the results illustrated in Fig. 12. However, note that in practice, the case where sources have small rates might not correspond to the regime of rare negative feedback, which is assumed in our paper. A verification by simulation is thus required, and remains to be done.

### B. TCP throughput-loss formula

In the steady-state, supposing that $g_l(\cdot) << 1$, $q_i \triangleq \sum_{l\in\mathcal{L}} A_{l,i}\, g_l(f_l)$ corresponds to the packet loss-ratio. In the limit case $\dot{x}^* = 0$, as $t \to \infty$, from the limit mean ODE (20) one can obtain

$$x_i^* = \sqrt{\frac{r_i}{\eta_i \tau_i q_i^*}}, \text{ as } q_i^* \to 0. \tag{35}$$

For a TCP connection, (35) is referred to as a TCP throughput-loss formula

$$x_i^* = \frac{C}{\tau_i\sqrt{q_i^*}}\ \mathrm{pck/s}, \tag{36}$$

where, from (35), $C = \sqrt{2}$ (resp. $C = 1$), for not delayed ACKs $r_i = 1/\tau_i$ (resp. for delayed ACKs $r_i = 0.5/\tau_i$), and in both cases $\eta_i = 1/2$. The reported measurements [2] indicate that $C$ varies substantially for different TCP implementations.

In the derivation of (35) we used a conditional probability of negative feedback equal to $x_i\tau_i \sum_{l:A_{l,i}>0} g_l(f_l)$. However, in Section II-B it is shown that, in general, conditional probability

TABLE II

LSE CORRECTION OF THE TCP THROUGHPUT-LOSS FORMULA (over $q_i^* \in [10^{-4}, 10^{-1}]$ for $\tau_i \in$ [10 ms 600 ms]).

| ACK Type | LSE($\hat{C}$), $\alpha = 0.5$ | LSE($\hat{C}$, $\alpha$) |
|---|---|---|
| not delayed ACK | 1.2997 | 1.1863, 0.5274 |
| delayed ACK | 0.9377 | 0.8731, 0.5215 |

of negative feedback has the form $1-\prod_{l:A_{l,i}>0}[1-g_l(f_l)]^{x_i\tau_i}$, from which the former expression follows for $g_l(\cdot) \to 0$. In reference to [2], we consider a single bottleneck so that in the steady state $g_l \equiv q_i^*$ is the packet loss-ratio. Hence, using the exponential form of the probability of negative feedback in the limit mean ODE, $\dot{x^*} = 0$, as $t \to \infty$, yields a throughput-loss formula of the form

$$r_i - (r_i + \eta_i x_i^*)[1-(1-q_i^*)^{x_i^*\tau_i}] = 0. \tag{37}$$

Let $y_i^*$ be a solution of (37) and $x_i^*$ be given by (35). Then we compute a correction factor $C^* = y_i^* C/x_i^*$, which is plotted in Fig. 13. It is evident that $C^*$ depends significantly on the loss-ratio $q_i^*$. We computed Least Square Error (LSE) fitting of the function $x_i = \hat{C}/(\tau_i q_i^\alpha)$ to the solution of (37), in respect to parameters ($\hat{C}$, $\alpha = 0.5$), and ($\hat{C}$, $\alpha$), over values of $q_i^*$ and $\tau_i$ as indicated in Table II. Finally, it turns out that the fitted values (Table II) are in agreement with the parameters fitted to the measured data of a set of TCP implementations [2].

## VI. CONCLUSION

We have shown that under certain reasonable assumptions, congestion avoidance based on additive-increase and multiplicative-decrease shares throughput according to a generalization of $F_A$ fairness, which explicitly accounts for communication delays and differences in round-trip times. It is important to note that this influence does not come into play because of the delayed feedback to the end-systems, but it is a fundamental consequence of the diversity of the rate updating intervals. It is demonstrated that different fairness objectives can be achieved depending on the values of additive increments and RTTs. It is shown that throughput results [1] are an asymptotic case of $F_A^h$, for small additive-increase/multiplicative-decrease ratio relatively to connection throughput. We also gave some insight into the bias against flows with long round-trip times. Further research should concentrate on analyzing the impact of RTT variations due to queueing.

## ACKNOWLEDGMENT

The authors would like to thank Vivek Borkar for his helpful comments on stochastic approximation algorithms.

## REFERENCES

[1] Sally Floyd, "Connections with multiple congested gateways in packet-switched networks part 1: One-way traffic," *Computer Communication Review (also available from: ftp://ftp.ee.lbl.gov/papers/gates1.ps.Z)*, vol. 21, no. 5, pp. 30–47, October 1991.

[2] Matthew Mathis, Jeffrey Semke, Jamshid Mahdavi, and Teunis Ott, "The macroscopic behavior of the tcp congestion avoidance algorithm," *Computer Communication Review*, vol. 27, no. 3, July 1997.

[3] Van Jacobson and Michael J. Karels, "Congestion avoidance and control," in *Proc. of the ACM SIGCOMM'88*, Stanford, August 1988, pp. 314–329.

[4] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Computer Networks and ISDN Systems*, vol. 17, pp. 1–14, June 1989.

[5] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, 1998.

[6] Paul Hurley, Jean-Yves Le Boudec, and Patrick Thiran, "A note on the fairness of additive increase and multiplicative decrease," in *Proceedings of ITC-16 (also available from: http://icawww.epfl.ch)*, Edinburgh, UK, June 1999.

[7] T. V. Lakshman and U. Madhow, "The performance of tcp for networks with high bandwidth–delay products and random loss," *IEEE/ACM Transactions on Networking*, vol. (5)3, pp. 336–350, June 1997.

[8] Harold J. Kushner and G. George Yin, *Stochastic Approximations Algorithms and Applications*, Springer-Verlag, 1997.

[9] B. Braden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Patridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski, and L. Zhang, "Recommendations on queue management and congestion avoidance in the internet," *IETF RFC 2309*, April 1998.

[10] Sally Floyd and Van Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. on Networking*, vol. 1, no. 4, pp. 397–413, August 1993.

[11] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms," in *Proc. of the IEEE INFOCOM'99*, March 1999, pp. 1395–1403.

[12] Thomas R. Henderson, Emile Sahouria, Steven McCanne, and Randy H. Katz, "On improving the fairness of tcp congestion avoidance," in *Proc. of the IEEE GLOBECOM '98*, Sydney, Australia, November 1998, IEEE.

[13] Lennart Ljung and Torsten Söderström, *Theory and Practice of Recursive Identification*, The MIT Press, Cambridge, Massachusetts, London, England, 1983.

[14] Albert Benveniste, Michel Métivier, and Pierre Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, 1987.

## APPENDIX

## I. THE LIMIT MEAN ODE

*Theorem 1* ([8], Ch. 12.3, p. 364–365) Assume (12)-(16). Then

$$\{x_i^\gamma(\cdot), \tau_i^\gamma(\cdot), \hat{x}^\gamma(\cdot), N_i^\gamma(\cdot), i \in \mathcal{S}\}$$

is tight in $D^{4|\mathcal{S}|}[0,\infty)$. Let $\gamma$ index a weakly convergent subsequence whose weak sense limit we denote by

$$(x_i(\cdot), \tau_i(\cdot), \hat{x}(\cdot), N_i(\cdot), i \in \mathcal{S}).$$

Then the limits are Lipschitz continuous with probability one and

$$x_i(t) = \hat{x}_i(\tau_i(t)),\ \hat{x}_i(t) = x_i(N_i(t)) \tag{38}$$

$$N_i(\tau_i(t)) = t. \tag{39}$$

Also,

$$\tau_i(t) = \int_0^t \bar{u}_i(\hat{x}(\tau_i(s)))ds, \tag{40}$$

$$\dot{x}_i(t) = \bar{h}_i(\hat{x}(\tau_i(t))) + z_i(t), \tag{41}$$

$$\dot{\hat{x}}_i = \frac{\bar{h}_i(\hat{x})}{\bar{u}_i(\hat{x})} + \hat{z}_i,\ i \in \mathcal{S}, \tag{42}$$

where the $z_i$ and $\hat{z}_i$ serve the purpose of keeping the paths in the interval $[a_i, b_i]$.

If $L_C$ is asymptotically stable in the sense of Lyapunov, then $\forall \mu > 0\ \exists T_\mu > 0$ s.t. for $t \geq T_\mu$, $||\hat{x}(t), L_C|| < \mu$. Also, $\forall T > T_\mu$,

$$\limsup_\gamma P\left\{\sup_{T_\mu \leq t \leq T} ||\hat{x}^\gamma(t), L_C|| \geq \mu\right\} = 0. \tag{43}$$

For large $T$ and $T_1 > T$, $\hat{x}^\gamma(\cdot)$, $t \in [T, T_1]$, spends nearly all of its time in a small neighborhood of $L_C$. Now, drop the constraint set $C$ and suppose that $\{\hat{x}_n^\gamma; \gamma, n\}$ is bounded with probability one. Then the above conclusions continue to hold with $z_i(t) = \hat{z}_i(t) = 0$ and $L_C$ replaced by some limit set of (42).