Fair End-to-End Window-Based Congestion Control

Jeonghoon Mo and Jean Walrand, *Fellow, IEEE*

***Abstract*—In this paper, we demonstrate the existence of fair end-to-end window-based congestion control protocols for packet-switched networks with first come–first served routers. Our definition of fairness generalizes proportional fairness and includes arbitrarily close approximations of max–min fairness. The protocols use only information that is available to end hosts and are designed to converge reasonably fast.**

Our study is based on a multiclass fluid model of the network. The convergence of the protocols is proved using a Lyapunov function. The technical challenge is in the practical implementation of the protocols.

***Index Terms*—Bandwidth sharing, congestion control, fairness, TCP, window.**

I. Introduction

WE STUDY the existence of fair end-to-end congestion control schemes. More precisely, the question is that of the existence of congestion control protocols that converge to a fair equilibrium without the help of the internal network nodes, or routers. Using such a protocol, end-nodes, or hosts, monitor their connections. By so doing, the hosts get implicit feedback from the network such as round-trip delays and throughput but no explicit signals from the network routers. The hosts implement a window congestion control mechanism. Such end-to-end control schemes do not need any network configuration and therefore could be implemented in the Internet without modifying the existing routers or the IP protocol.

The Internet congestion control is implemented in end-to-end protocols. The motivation for such protocols is that they place the complex functions in the hosts and not inside the network. Consequently, only the hosts that want to implement different complex functions need to have their software upgraded. Another justification, which is more difficult to make precise, is that by keeping the network simple it can scale more easily.

TCP is the most widely used end-to-end protocol in the Internet. When using TCP [12], a source host adjusts its windows size, the maximum amount of outstanding packets it can send to the network, to avoid overloading routers in the network and the destination host.

Many researchers have observed that, when using TCP, connections with a long round-trip time that go through many bottlenecks have a smaller transmission rate than the other connections [4], [6], [22]. To improve fairness, Floyd and Jacobson [5] proposed a "constant rate adjustment" algorithm and Handerson *et al.* [9] simulated a variation of this scheme. However, choosing the parameters of such algorithms is still an open problem.

Thus, although end-to-end protocols such as those implemented in TCP are very desirable for extensibility and scalability reasons, they are unfair. Roughly, a fair scheme is one that does not penalize some users arbitrarily. Accordingly, the question that arises naturally is the existence of fair end-to-end congestion protocols.

In an early paper, Jaffe [14] shows that power, defined as $(\text{throughput})/(\text{average delay})$, cannot be optimized in a distributed manner.

Chiu and Jain [3] show that in a network with N users that share a unique bottleneck node, a linear increase and multiplicative decrease algorithm converges to an efficient and fair equilibrium. Most current implementations of TCP window-based control use a linear increase and multiplicative decrease of the window size, as suggested in [12]. However, these implementations control the size of their window and not their transmission rate. Moreover, simple examples show that the algorithm does not converge to an efficient and fair equilibrium in networks with multiple bottleneck nodes.

Shenker [28] considers a limited class of protocols and argues that "no aggregate feedback control is guaranteed fair." This statement suggests that end-to-end control cannot guarantee convergence to a fair equilibrium. Unfortunately, the class of protocols that he considers excludes many implementable end-to-end protocols. Jain and Charny refers to [28] to justify the necessity of switch-based control for fairness [15], [2].

Recently, Kelly *et al.* [18] proposed *proportional fairness* and exhibited an aggregate feedback algorithm that converges to the point. In their scheme, each user is implementing a linear increase and multiplicative decrease of its rate based on an additive feedback from the routers the connection goes through. This protocol requires that the routers can signal the difference between their load and their capacity. In our protocol, each host controls its window size not its rate, based on the total delay. Our protocol can be viewed as a refinement of TCP congestion control algorithms.

Le Boudec *et al.* [10], [21] studied the exact form of fairness founded by the linear increase, multiplicative decrease algorithms. Massoulie and Roberts have done interesting work on bandwidth sharing which is similar to ours [23]. They proposed fixed window algorithms to achieve fairness.

In this paper, we revisit the fundamental question of the existence of fair end-to-end protocols and we provide a positive answer by constructing explicitly such protocols. The rest of the paper is organized as follows. In Section II, we present a multiclass fluid model for window-based control and theoretical re-

Manuscript received November 10, 1998; revised January 19, 1999 and October 4, 1999; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Floyd. This work was supported in part by a grant from the National Science Foundation.

J. Mo is with AT&T Labs, Middletown, NJ 07748 USA (e-mail: jhmo@att.com).

J. Walrand is with the University of California, Berkeley, CA 94704 USA.

Publisher Item Identifier S 1063-6692(00)09125-1.

sults about the model. Section III presents our generalized definition of fairness and its relation to window-based congestion control. In Section IV, we propose window-based end-to-end fair algorithms and prove their convergence to a fair equilibrium. Section V contains simulation results of our algorithms, which support our claims. Finally, Section VI draws conclusions and discusses future research directions.

II. Network Model

This section explains a multiclass network model and a mapping between window size vector and the rate vector defined by the model. Since it is through the rates that fair sharing is defined and the protocol can control the window, the relation between the rate vector and the window size vector is important. We show that given the window size vector and network topology, the rate vector is well defined.

A. Multiclass Fluid Model

We consider a *closed multiclass fluid network* with M links and N connections. The sender of connection $i \in \mathcal{N}$, where $\mathcal{N}$ is the set of users with the cardinality N, exercises a window-type flow control and adjusts the window size w_i of the connection. A connection follows a route that is a set of links. Link $j \in \mathcal{L}$ has capacity, or transmission rate, c_j, where $\mathcal{L}$ is the set of links with the cardinality M. We define the matrix $A = (A_{ij},\, i \in \mathcal{N},\, j \in \mathcal{L})$ where $A_{ij} = 1$ if connection i uses link j and $A_{ij} = 0$, otherwise. Let also $A_{i\cdot} := \{j | A_{ij} = 1\}$ be the set of links that connection i uses and $A_{\cdot j} := \{i | A_{ij} = 1\}$ the set of connections that use link j.

Each connection i has a fixed round-trip propagation delay d_i, which is the minimum delay between the sending of a packet by the sender host and the reception of its acknowledgment by the same host. We assume that the processing times are negligible. A typical acknowledgment delay comprises d_i and some additional queueing delay in bottleneck routers. Let x_i be the flow rate of the ith connection for $i \in \mathcal{N}$. For $j \in \mathcal{L}$, we assume that every link j has an infinite buffer space, and we designate by q_j the work to be done by link j. By definition, q_j is the ratio of the queue size in the buffer of link j divided by the capacity c_j. The service discipline of the links is first come–first served.

We consider a fluid model of the network where the packets are infinitely divisible and small. This model is represented by following equations:

$$A^T x - c \le 0 \tag{1}$$

$$Q(A^T x - c) = 0 \tag{2}$$

$$X(Aq + d) = w \tag{3}$$

$$x \ge 0, \quad q \ge 0 \tag{4}$$

where

$$x = (x_1, \ldots, x_N)^T$$
$$c = (c_1, \ldots, c_M)^T$$
$$q = (q_1, \ldots, q_M)^T$$
$$d = (d_1, \ldots, d_N)^T$$
$$X = \mathrm{diag}(x)$$
$$Q = \mathrm{diag}(q).$$

The inequality (1) expresses the capacity constraints: the sum of the rates of flows that go through a link cannot exceed the capacity of the link. The identity (2) can be written as

$$q_j \left[(A^T x)_j - c_j\right] = 0, \quad \text{for } j \in \mathcal{L}.$$

The jth identity means that if the rate $(A^T x)_j$ through link j is less than the capacity c_j of the link, then the queue size q_j at that link is equal to 0. Finally, the identity (3) which can be written as

$$x_i \left[(Aq)_i + d_i\right] = w_i, \quad i \in \mathcal{N}$$

means that the total number of packets w_i for each connection $i (\in \mathcal{N})$ is equal to the number $x_i d_i$ of packets in transit in the transmission lines plus the total number $x_i (Aq)_i$ of packets of connection i stored in buffers along the route. To clarify the meaning of $x_i (Aq)_i$, note that

$$x_i (Aq)_i = x_i \sum_j A_{ij} q_j = \sum_j A_{ij} x_i q_j.$$

Now, $c_j q_j$ is the number of packets in the buffer of link j and a fraction x_i / c_j of these packets are of connection i. Thus, $(c_j q_j)(x_i / c_j) = x_i q_j$ is the backlog of packets of connection i in the buffer of link j. Summing over all j such that connection i goes through link j shows that $x_i (Aq)_i$ is the total backlog of packets of connection i.

Note that our model assumes that, for each link j, the contribution to the queue size of connection i is proportional to its flow rate x_i. This assumption is consistent with the fluid assumption under which the packets are infinitely divisible.

We rewrite the identity (3) as follows:

$$x_i = \frac{w_i}{D_i} \quad \text{where } D_i = d_i + (Aq)_i, \quad \text{for } i \in \mathcal{N}. \tag{5}$$

The identity (5) means that the flow rate x_i of connection i is equal to the ratio of the window size w_i of the connection divided by its total round-trip delay D_i. The total delay D_i consists of fixed propagation delay d_i plus a variable queueing delay which depends on congestion in the network. Accordingly, the flow rate x_i of connection i is a function of not only the window size w_i of the connection but also of the window sizes of the other connections. When the network is not congested, $q = (q_1, \ldots, q_M) = 0$ and the flow rates are proportional to the window sizes. However, as congestion builds up, $q \neq 0$ and the rates are no longer linear in the window sizes.

B. Mapping from Window Size Vectors to Rate Vectors

In this subsection we prove that the flow rate vector x is a well-defined function of the window size vector w and we derive some properties of the function. This result is intuitively clear and its proof is a confirmation that the model captures the essence of the physical system. Before proving the uniqueness of the rate vector x, we first show the existence of a rate vector x that solves the relations that characterize the fluid model.

Theorem 1: For given values of (w, A, d, c), there exists at least one rate vector x which satisfies the relations (1)–(4).

Proof: Let $U = \sum_{i=1}^{N} w_i$. Then $q_j \in [0, U]$ for $j = \in \mathcal{L}$. For $q \in [0, U]^M$, let

$$x_i(q) := \frac{w_i}{d_i + (Aq)_i}, \quad i \in \mathcal{N}$$
$$f(q) := c - A^T x(q)$$

and

$$h_j(q) = -f_j^2(q), \quad j \in \mathcal{L}.$$

Fix $j \in \{1, \ldots, M\}$ and $q^j := (q_1, \ldots, q_{j-1}, q_{j+1}, \ldots, q_M)$ in $[0, U]^{M-1}$. We claim that $h_j(q) = h_j(q_1, \ldots, q_M)$ is a quasi-concave function of q_j. By definition of quasi-concavity, this means that $\{q_j | h_j(q) \geq a\}$ is convex for all $a \in \mathcal{R}$. To verify the claim, note that

$$f_j(q_j, q^j) = c_j - \sum_i A_{ji} \frac{w_i}{d_i + (Aq)_i} = c_j - \sum_i A_{ji} \frac{w_i}{d_i + \sum_{l \neq j} A_{il} q_l + A_{ij} q_j}$$

is increasing and concave on $[0, U]$. Indeed, f_j is the sum of increasing concave functions on $[0, U]$. If $f_j(0, q^j) \geq 0$, then h_j is a decreasing function, which is quasi-concave on $[0, U]$. Also, in that case, $\arg\max_{q_j} h_j(q_j, q^j) = 0$. On the other hand, if $f_j(q^j) < 0$ then h_j is a unimodular function which increases on $[0, q_j^*]$ and decreases on $(q_j^*, U]$, which also is quasi-concave. Moreover, in that case, $\max_{q_j} f_j(q_j, q^j) = 0$. This proves the quasi-concavity of h_j.

By the theorem of Nash [11], the quasi-concavity of $h_j(q_j, q^j)$ in q_j for any fixed q^j implies that there exists at least one vector $q^* \in [0, U]^M$ such that

$$q_j^* = \arg \max_{q_j \in [0, U]} h_j(q_1^*, \ldots, q_{j-1}^*, q_j, q_{j+1}^*, \ldots, q_M^*) \quad \text{for } j \in \mathcal{L}.$$

Let q^* be such that (6) holds and let $x^* = x(q^*)$. We claim that x^* is a solution of (1)–(4). To verify the claim, observe that our proof of the quasi-concavity shows that either $q_j^* = 0$ or $f_j(q^*) = 0$ and that in both cases $f_j(q^*) \geq 0$. Hence, $q_j^* f_j(q^*) = 0$ and (2) follows. Moreover, $f_j(q_j^*, q^j) \geq 0$ is equivalent to (1). Additionally, (3) and (4) are trivial by construction. ■

Theorem 1 guarantees that the existence of a rate vector and the next theorem proves its uniqueness.

Theorem 2: Given (w, A, d, c), the flow rate vector x satisfying (1)–(4) is unique.

For readability, we briefly sketch the proof here. Refer to Appendix A for the details of the proof.

Sketch of the Proof: The proof is composed of two parts. In the first part, we show that given (w, A, d, c), the set of bottleneck links B is uniquely determined. We assume two sets of bottleneck B_1 and B_2 and derive a contradiction using Farkas' lemma (see, e.g., [24]). In the second part, we show that, given (w, A, d, c) and B, the vector of flow rates x is unique. To prove the uniqueness, we show that the partial derivative matrix of the fixed point equation (40) is positive definite and use the partial results of Rosen [27]. Combining those two parts completes the proof of the uniqueness.

C. Comments on the Mapping

1) Queue Size is Not Unique: Although the rate vector is uniquely determined from the window sizes, the workload vector q generally is not, as the following example shows. Consider a network with two bottleneck links in series with the same capacity c and a single connection with window size w. If $(w/d) > c$, then the queues build up in the links. For this network, any vector (q_1, q_2) such that $q_1 + q_2 = (w/c) - d$ is a solution of (1)–(4).

The following corollary shows a sufficient condition for q to be determined uniquely.

Corollary 1: If $\mathrm{rank}(A_B)$ is equal to the number $|B|$ of bottleneck links, then (w, A, c, d), uniquely determines the vector q.

Proof: From the uniqueness of flow rate vector x and (3), $q = (A_B^T A_B)^{-1} A_B^T (X_B^{-1} w_B - d_B)$. The inverse exists from the full rank assumption.

2) Bottleneck: The following lemma provides sufficient conditions for links not to be bottlenecks.

Lemma 1: For any given window size vector w, 0–1 matrix A, and diagonal matrix D

1) if ${A_{\cdot j}}^T D^{-1} w \leq c_j$, then $q_j = 0$;
2) $A^T D^{-1} w \leq c$ if and only if $q = 0$.

Proof:

1) Assume that $q_j > 0$. This implies $x_i < (w_i/d_i)$. Now, $A_{\cdot j}^T D^{-1} w \leq c_j$ implies $A_{\cdot j}^T x < c_j$, since $D^{-1} w$ is the upper bound on x. From (2), $q_j = 0$, which is a contradiction. Hence, $q_j = 0$.
2) If $q = 0$, the window size vector $w = Xd$ from (3) where $D = \mathrm{diag}(d_i, i = 1, \ldots, N)$. By (1), we prove if part. The only if part is obvious from part 1). ■

The converse of part 1) is not always true, as can be seen from the next example. Let $M = 2$, $N = 2$, $C = (5, 5)^T$, $d = (1, 1)^T$, and

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

If $w = (10, 20)$, clearly, $q_2 = 0$, the flow rate out of resource $1 \leq 5$, but $A_{\cdot 2}^T D^{-1} w = 10 > 5$.

3) Some Properties of the Mapping: Let F: $W \rightarrow X$ be the mapping from the window space W to a flow rate space X defined by (1)–(4). Let $B(w)$ be the set of bottlenecks for the window sizes w. We call w an interior point if there is $\epsilon > 0$ such that $B(\overline{w})$ are same for all $\overline{w} \in$ neighborhood $N_\epsilon(w)$ of w. Otherwise, w is said to be a boundary point.

F is a continuous function but is not differentiable at the boundary point. We illustrate this by the next example. For complete proofs, see Appendix B.

Consider the network and connections in Fig. 1(a). Two users are sharing one link and each uses another link. Fig. 1(b) is a plot of x_1 along the horizontal dotted line 1 in Fig. 1(c). Fig. 1(b) shows that x_1 is a continuous nondecreasing function of the window size w_1, but is not differentiable at the points where

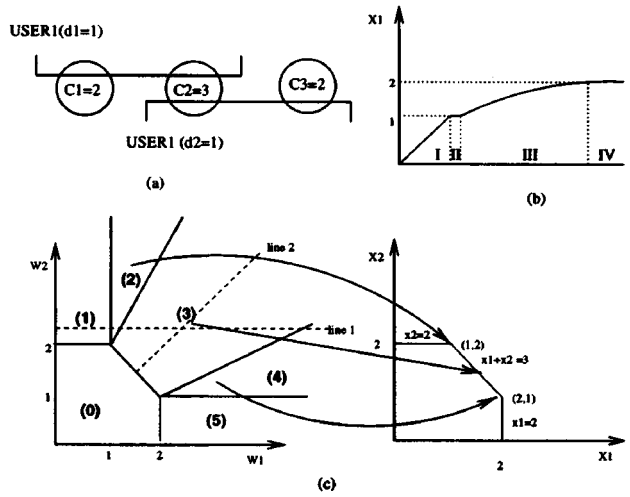


Fig. 1. (a) Network topology. (b) Flow rate x versus window size. (c) Mapping between x and w.

the set of bottlenecks changes. Each region I, II, III, and IV corresponds to different sets of bottlenecks. For example, in region I, user 1 does not suffer from any bottlenecks, but user 2 does.

Fig. 1(c) shows the mapping $x = F(w)$. If $w \in (0)$, the queues are empty, and w and x are such that $w_i = x_i d_i$, so that $x = F(w)$ is differentiable in that region. If $w \notin (0)$, F is no longer one to one. For instance, $F(w) = (1, 2)$ for all $w \in (2)$ and $F(w) = (2, 1)$ for all $w \in (4)$.

Let $F^{-1}(x) = \{w | F(w) = x\}$. The dimension of $F^{-1}(x)$ is related to the number of bottlenecks. To be precise, the dimension of $F^{-1}(x)$ is same as the rank of A_B. This property follows from $w = Xd + XAq$. Since $XAq = XA_Bq_B$, $F^{-1}(w)$ is a positive cone of XA_B with vertex Xd, as we now illustrate in Fig. 1(c). When $q = 0$, the inverse image of F is a point, of which the dimension is 0. When $x = (1.5, 1.5)$, $F^{-1}(x)$ is the dotted line 2 in the figure, whose dimension is 1. When $x = (2, 1)$ or $(1, 2)$, the number of bottlenecks is 2, which is the dimension of $F^{-1}(x)$.

III. Fairness

Fairness has been defined in a number of different ways so far. The notion of fairness characterizes how competing users should share the bottleneck resources. In this section, we review and compare standard definitions of fairness and generalize them.

A. Max–Min Fairness

One of the most common fairness definitions is *max–min* or *bottleneck optimality* criterion [13], [1], [8], [16], [2]. A feasible flow rate x is defined to be *max–min fair* if any rate x_i cannot be increased without decreasing some x_j which is smaller than or equal to x_i [1]. Many researcher have developed algorithms achieving max–min fair rates [1], [16], [2]. But a max–min fair vector needs global information [25], and most of those algorithms require exchange of information between networks and hosts. In [8], Hahne suggested a simple round-robin way of control, but it requires all the links perform round-robin scheduling and it needs to be guaranteed that packets of users are ready for all links.

B. Proportional Fairness

Kelly [17] proposed *proportional fairness*. A vector of rates x^* is proportionally fair if it is feasible, that is, $x^* \geq 0$ and $A^T x^* \leq c$, and if for any other feasible vector x, the aggregate of proportional change is negative.[1]

$$\sum_i \frac{x_i - x_i^*}{x_i^*} \leq 0. \tag{6}$$

In [18], Kelly *et al.* suggested a simple algorithm that converges to the proportionally fair rate vector.[2]

Proportional fairness can be motivated in another way. Consider the following optimization problem (P):

maximize

$$g = \sum_i p_i f(x_i) \tag{7}$$

subject to

$$A^T x \leq c \tag{8}$$

over

$$x \geq 0 \tag{9}$$

where f is an increasing strictly concave function and the p_i are positive numbers.

Since the objective function (7) is strictly concave and the feasible region (8), (9) is compact, the optimal solution of (P) exists and is unique. Let $L(x, \mu) = g(x) + \mu^T(c - A^T x)$. The Kuhn–Tucker conditions [24] for a solution x^* of (P) are

$$\nabla g^T - \mu^T A^T = 0 \tag{10}$$

$$\mu_j(c_j - A_j^T x^*) = 0 \quad \text{for } j \in \mathcal{L} \tag{11}$$

$$A^T x^* \leq c \tag{12}$$

$$x^* \geq 0, \quad \mu \geq 0 \tag{13}$$

where $\nabla g^T = (p_1 f'(x_1), \ldots, p_n f'(x_n))$. When there is only one link and N connections, the optimal solution of (P) is $x_i = c/N$ for all i: all the connections have an equal share of the bottleneck capacity, irrespective of the increasing concave f. Indeed, (10) implies $f'(x_i) = \mu$ for all i, so that $x_i = f'^{-1}(\mu)$ for all i. If x is a proportionally fair vector then it solves (P) when $f(x) = \log x$ with $p_i = 1$ for all i. Thus, a proportionally fair vector is one that maximizes the sum of all the logarithmic utility functions.

The fair rate vector is not so simple when there are multiple bottlenecks. Consider the following network with two different bottlenecks and three connections. The max–min fair rate vector of this network is $(c_1/2, c_1/2, c_2 - (c_1/2))$ if $c_1 < c_2$, while the proportionally fair rate vector is not the same as the max–min

[1]Refer to http://www.ucl.ac.uk/uceemdb/work/phd.html for the discussion on why the inequality should not be strict.

[2]In [10], Hurley *et al.* showed that Kelly's algorithm is based on the assumption that the feedback is independent of the sending rate of the user. In the case of proportional feedback, i.e., when senders receive feedback proportional to their sending rates, they showed that the additive increase and multiplicative decrease algorithm does not lead to the proportional fair rate, but to some variant of it.

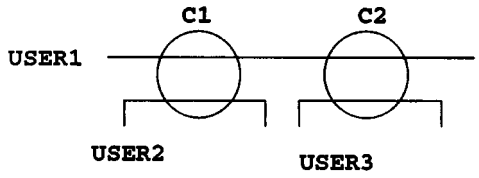


Fig. 2. Network with multiple bottlenecks.

fair rate in this case, since by decreasing the rate of user 1, the sum of the utility functions f increases. Hence, the optimal vector x depends on the utility function f when there are at least two bottlenecks.

It is the concavity of the function f that forces fairness between users. If f is a convex increasing function instead of concave, then to maximize the objective function g of (P), the larger flow rate x_i should be increased, since the rate of increase of $f(x_i)$ is increasing in x_i. When f is linear, the rate of increase of f is the same for all x. When f is concave, a smaller x_i favored, since $f'(x) > f'(y)$ if $x < y$.

C. (p, α)-Proportional Fairness

It is a matter of controversy what is a fair rate allocation for the network in Fig. 2. It can be argued that the max–min fair rate is desirable. On the other hand, connection 1 is using more resources than the others under the max–min fair rate. Generally, the problem is how to compromise between the fairness to users and the utilization of resources. The max–min definition gives the absolute priority to the fairness. The following definition is a generalization of proportional fairness and max–min fairness. When $\alpha = 1$, the given definition reduces to that of proportional fairness and as α becomes large, it converges to that of max–min.

Definition 1 [(p, α)-Proportionally Fair]: Let $p = (p_1, \ldots, p_N)$ and α be positive numbers. A vector of rates x^* is (α, p)-proportionally fair if it is feasible and for any other feasible vector x

$$\sum_i p_i \frac{x_i - x_i^*}{x_i^{*\alpha}} \leq 0. \tag{14}$$

Note that (14) reduces to (6) when $p = (1, \ldots, 1)^T$ and $\alpha = 1$.

The following lemma clarifies the relationship between the above definition and the problem (P).

Lemma 2: Define the function f_α as follows:

$$f_\alpha(x) := \begin{cases} \log x, & \text{if } \alpha = 1 \\ (1-\alpha)^{-1} x^{1-\alpha}, & \text{otherwise.} \end{cases}$$

Then the rate vector x^* solves the problem (P) with $f = f_\alpha$ if and only if x^* is (p, α)-proportionally fair.

Proof: Let x^* be a solution of (P). We show that x^* is (p, α)-proportionally fair. Multiplying (10) by $(x - x^*)$, we find $\nabla g^T(x - x^*) = \mu^T A^T(x - x^*)$. Summing the identity (11) over j, we obtain $\mu^T c = \mu^T A^T x^*$. Multiplying (8) by μ, we get $\mu^T A^T x \leq \mu^T c$. Combining these relations, we see that $\mu^T A^T x \leq \mu^T c = \mu^T A^T x^*$. Therefore, $\nabla g^T(x - x^*) = \mu^T A^T(x - x^*) \leq 0$. But $\nabla g^T(x - x^*) = \sum_i (p_i(x - x^*)/(x^{*^\alpha})$. Hence, $\sum_i (p_i(x - x^*)/(x^{*^\alpha})) \leq 0$. We have shown that the solution of (P) satisfies (14).

To prove the converse, assume that x^* is (p, α)-proportionally fair. By showing that it solves (P), we complete the proof. First note that $g(x) = g(x^*) + \nabla g(x^*)^T(x - x^*) + (1/2)(x - x^*)^T \nabla^2 g(x^*)(x - x^*) + o(\|x - x^*\|^2)$. The second term $\nabla g(x^*)^T(x - x^*) = \sum_i (p_i(x - x^*)/(x^{*^\alpha})) \leq 0$ for all feasible x and the third term is strictly negative by negative definiteness of $\nabla^2 g$. Hence, x^* is a local minimum. Since P has a unique global solution, x^* solves (P). ■

The next lemma explains the relationship between max–min fair rate and the parameter α.

Lemma 3: If $h(x)$ is a differentiable increasing concave negative function when $x \geq 0$, the solution of (P) with $f_\alpha = -(-h)^\alpha$ approaches the max–min fair rate vector as $\alpha \to \infty$.

Proof: We will consider the case in which $p_i = 1$ for all i. Extension to the general case is straightforward. Let x^α be the optimal solution of (P) with f_α and $\mathcal{X} = \{x | A^T x \leq c, x \geq 0\}$. Since $\{x^\alpha\}$ is a sequence in a compact set $\mathcal{X}$, there exists a subsequence, say $\{\alpha_k, k \geq 1\}$, of α such that x^{α_k} converges to some $\overline{x} \in \mathcal{X}$ as $k \to \infty$.

We show that $\overline{x}$ is the max–min vector. Since that max–min vector is unique, this will prove that all the limit points of $\{x^\alpha\}$ are that unique max–min vector and, therefore, that $\{x^\alpha\}$ converges to the max–min vector, as we want to show.

Assume that $\overline{x}$ is not a max–min vector. Then there exists a user i whose rate $\overline{x}_i$ can be increased with decreasing the rates of other users $\overline{x}_j$ which are greater than $\overline{x}_i$. Let L_1 be the set of saturated links used by i and L_2 the set of the other links used by i. For each link $l \in L_1$, there exists a user, say $u(l)$, whose rate $\overline{x}_{u(l)}$ is greater than $\overline{x}_i$, that is, $\overline{x}_{u(l)} > \overline{x}_i$ for $j \in L_1$. Define δ by

$$\delta = \tfrac{1}{5} \min \left\{ \min_{l \in L_1} (\overline{x}_{u(l)} - \overline{x}_i), \min_{l \in L_2} (c_l - (A^T \overline{x})_l) \right\}.$$

For simplicity, we denote f_{α_k} and x^{α_k} by f_k and x^k. From the convergence of x^k to $\overline{x}$, we can find k_0 such that for all $k \geq k_0$, for all j

$$\overline{x}_j - \frac{\delta}{N} \leq x_j^k \leq \overline{x}_j + \frac{\delta}{N} \tag{15}$$

where N is the number of users. Define sequence of vectors y^k as follows:

$$y_j^k = \begin{cases} x_j^k + \delta, & \text{if } j = i \\ x_j^k - \delta, & \text{if } j = u(l) \text{ for } l \in L_1 \\ x_j^k, & \text{otherwise.} \end{cases} \tag{16}$$

It can be shown $y^k \geq 0$ and $A^T y^k \leq c$ for $k \geq k_0$ without difficulty since we choose δ small enough. We now establish a contradiction with the optimality of x^k. Consider the expression A_k defined by

$$A_k = \sum_j \left(f_k(y_j^k) - f_k(x_j^k) \right).$$

From the optimality of x^k, we have $A_k \leq 0$. Now

$$A_k = f_k(x_i^k+\delta)-f_k(x_i^k)+\sum_{l\in L_1}\left(f_k(x_{u(l)}^k-\delta)-f_k(x_{u(l)}^k)\right).$$

From the theorem of intermediate values, there exist numbers c_i^k such that

$$\begin{cases} x_i^k \leq c_i^k \leq x_i^k+\delta \\ f_k(x_i^k+\delta)-f_k(x_i^k) = f_k'(c_i^k)\delta. \end{cases}$$

Combining with (15), we find $c_i^k \leq \overline{x}_i+(\delta/N)+\delta \leq \overline{x}_i+2\delta$. Similarly, there exist some numbers $c_{u(l)}^k$ such that

$$\begin{cases} x_{u(l)}^k-\delta \leq c_{u(l)}^k \leq x_{u(l)}^k \\ f_k(x_{u(l)}^k-\delta)-f_k(x_{u(l)}^k) = -f_k'(c_{u(l)}^k)\delta \end{cases}$$

and combining with (15) we find also $c_{u(l)}^k \geq \overline{x}_i+3\delta$. Thus

$$\begin{aligned} A_k &= \delta\left(f_k'(c_i^k)-\sum_{l\in L_1}f_k'(c_{u(l)}^k)\right) \\ &\geq \delta\left(f_k'(\overline{x}_i+2\delta)-Gf_k'(\overline{x}_i+3\delta)\right) \\ &= \delta f_k'(\overline{x}_i+2\delta)\left(1-G\frac{f_k'(\overline{x}_i+2\delta)}{f_k'(\overline{x}_i+3\delta)}\right) \end{aligned}$$

where G is the cardinality of L_1. The inequality follows from the concavity of f_k and and the bounds on c_i^k and $c_{u(l)}^k$. Since the last term in the parenthesis tends to 1 as k increases and $f_k' > 0$, for k large enough $A_k > 0$, which is a contradiction. ■

Corollary 2: The (p,α)-proportionally fair rate vector approaches the max–min fair rate vector as $\alpha \to \infty$.

This result follows from Lemma 3, since $f_\alpha(x) = (-1/(\alpha-1))(1/x)^{\alpha-1}$ in Lemma 2 satisfies the conditions of Lemma 3 with $h = -(1/x)$. Note that the constant $(\alpha-1)$ does not affect the solution of (P).

IV. Fair End-to-End Algorithms

In this section, we propose protocols that achieve fairness as defined in the previous section. We first show that the backlog of each user can play the role of a decoupled fairness criterion and prove the convergence of the proposed algorithms.

One implicit assumption for the convergence proof is that the end user knows network conditions immediately, which is not necessarily true. We do not consider feedback delays in the convergence proof. However, in order to support our claims, we include simulation results. The impact of the delay on the convergence will be a topic of further studies.

A. Decoupled Fairness Criteria

One of the major difficulties in achieving end-to-end fairness is that it is hard for end users to know the fair share of the network, since the fair share is a function of not only other users but also of the topology of the network. When n users are sharing a single bottleneck, the fair share will be $1/n$ capacity. To know the fair share, each end user should know the behavior of other users, which is not possible. Researchers have used similar arguments to claim the impossibility of achieving end-to-end fairness. This section proposes a decoupled fairness criteria, which each user can use to achieve fairness without considering the behavior of other users.

Let w_i, x_i, and d_i designate the window size, the rate, and the round-trip propagation time of connection i, respectively. Let $p_i > 0$ for $i \in \mathcal{N}$. Define

$$s_i = w_i - x_i d_i - p_i, \quad \text{for } i \in \mathcal{N}. \tag{17}$$

The expression $w_i - x_i d_i$ can be interpreted as backlog of user i and p_i as a target value. Hence, s_i is the difference between the actual and the target backlog. The next theorem shows that any window vector w such that $s_i = 0$ for all i corresponds to a $(p, 1)$-proportionally fair rate vector x.

Theorem 3: There is a unique window vector w such that $s_i = 0$ for $i \in \mathcal{N}$. Moreover, the corresponding rate vector $x(w)$ defined by (1)–(4) is a $(p, 1)$-proportionally fair rate vector.

The theorem states that the backlog of user i should be p_i for the rates to be proportionally fair. The interpretation of the theorem is that when an estimated queue size $w_i - x_i d_i$ is the same as the target backlog p_i of connection i, for all users, the resulting rate vector is proportionally fair. Note that all the parameters in the equation of Theorem 5 are observable or, at least, can be estimated by an end-user. The window size w_i is controlled by the end-user, and the flow rate x_i can be determined by observing return acknowledgments.

Proof: For any given $w \in [0, \infty)^N$, let x be the rate vector that corresponds to w, as defined by (1)–(4). Fix $w \in W$. From (3), we get $XAq = w - Xd = p$ where $p = (p_1, \ldots, p_N)^T$. The last equality follows from the definition of W. Hence, (10) is satisfied. From (1), (2), and (4), if we replace μ_j with q_j, the optimality conditions (11)–(13) of problem (P) hold for $f(x) = p_i \log x$ for $i \in \mathcal{N}$. By Lemma 2, x is a $(p, 1)$-proportionally fair rate vector. The uniqueness of w follows from the uniqueness of the $(p, 1)$-proportionally fair rate vector x, (3), and $XAq=p$. ■

Observe that the workload vector q is same as the optimal dual variables μ of the problem (P) when the network is in the state of $(p, 1)$-proportional fairness. This theorem implies that by controlling the total backlogs of the network, we can operate the network at the $(p, 1)$-proportionally fair point.

This theorem can be extended to the (p, α)-proportionally fair case. Let $p_i > 0$ for $i \in \mathcal{N}$ and $\alpha > 1$. Define

$$s_i^\alpha = w_i - x_i d_i - \frac{p_i}{x_i^{\alpha-1}}, \quad \text{for } i \in \mathcal{N}. \tag{18}$$

Theorem 4: There is a unique window vector w such that $s_i^\alpha = 0$ for all i. Moreover, the corresponding rate vector $x(w)$ defined by the identities (1)–(4) is a (p, α)-proportionally fair rate vector.

Proof: Note that if $s_i^\alpha = 0$, then $w_i - x_i d_i = (p_i/x^{\alpha-1}) = x_i A_i.q$. Consequently, $x^\alpha Aq = p$, which is the optimality condition (10). The other conditions, (12), (11), and (13), are satisfied, as can be shown as in the previous proof. ■

Note the difference that the target backlog $(p_i/x_i^{\alpha-1})$ is not constant but a decreasing function of rate x_i when $\alpha > 1$. Since the target backlog of a smaller rate connection is greater than that of larger rate connection, the smaller rate connection tries to put more backlogs in the network. By so doing, the equilibrium they achieve can be closer to fair rate.

B. $(p, 1)$-Proportionally Fair Algorithm

The previous section showed that $w_i - x_i d_i$ can be used as a decoupled fairness criteria. In this section, we construct an end-to-end window control that achieves convergence to the proportionally fair rate vector. Define $\overline{d}_i = d_i + A_i.q$ for $i \in \mathcal{N}$. That is, $\overline{d}_i$ is the measured round-trip delay of connection i. Fix $\kappa > 0$.

Consider the following system of differential equations:

$$\frac{d}{dt}w_i(t) = -\kappa \frac{d_i}{\overline{d}_i}\frac{s_i}{w_i} \tag{19}$$

$$s_i = w_i - x_i d_i - p_i \quad \text{for } i \in \mathcal{N}. \tag{20}$$

Theorem 5 states that if each user changes its window size based on (19) and (20), the rate vector achieved by the users will be $(p, 1)$-proportionally fair.

Theorem 5: Let $V(w) = \sum_{i=1}^{N}(s_i/w_i)^2$. Then V is a Lyapunov function for the system of differential equations (19), (20). The unique value minimizing V is a stable point of this system, to which all trajectories converge.

Define

$$J_x = \left[\frac{\partial x_i}{\partial w_j},\, i, j \in \mathcal{N}\right] \tag{21}$$

$$J_q = \left[\frac{\partial q_i}{\partial w_j},\, i \in \mathcal{L},\, j \in \mathcal{N}\right]. \tag{22}$$

Let B be the set of bottleneck links that correspond to w. Designate by A_B the submatrix of A obtained by keeping only the columns that correspond to a bottleneck link.

Lemma 4: The Jacobian $J_x = [(\partial x_i/\partial w_j),\, i, j \in \mathcal{N}]$ of $x(w)$ with respect to w is given by the following expression on the interior point:

$$J_x = \overline{D}^{-1}\left(I - XA_B(A_B^T X\overline{D}^{-1}A_B)^{-1}A_B^T\overline{D}^{-1}\right) \tag{23}$$

where $\overline{D} = \mathrm{diag}(d_i + A_i.q,\, i \in \mathcal{N})$, $X = \mathrm{diag}(x_i,\, i \in \mathcal{N})$.

For the proof of the Lemma 4, see Appendix B.

Proof of Theorem 5: We will first consider the case in which $w(t)$ is an interior point. Let $r_i = s_i/w_i$. Note that

$$\begin{aligned}\frac{d}{dt}V(w(t)) &= \sum_j \frac{\partial V}{\partial w_j}\cdot\frac{dw_j(t)}{dt}\\ &= -\sum_j\sum_i r_i\frac{dr_i}{dw_j}\dot{w}_j\\ &= -\kappa r^T J_r \dot{w}\end{aligned}$$

where $J_r := ((dr_i/dw_j),\, i \in \mathcal{N},\, j \in \mathcal{N})$ is the Jacobian of r. From $J_r = (XDW^{-2} + PW^{-2} - DW^{-1}J_x)$, $\dot{w} = D\overline{D}^{-1}r$, and (23)

$$\begin{aligned}&\frac{d}{dt}V(w(t))\\ &\quad= -\kappa r^T[(XDW^{-2} + PW^{-2} - DW^{-1}J_x)D\overline{D}^{-1}]\\ &\quad= -\kappa r^T[PW^{-2}D\overline{D}^{-1}\\ &\qquad + D\overline{D}^{-2}A_B(A_B^T X\overline{D}^{-1}A_B)^{-1}A_B^T\overline{D}^{-2}D]r.\end{aligned}$$

Identity (24) is obtained from identity (23). Note that the matrix in the bracket of (24) is positive definite. Hence $V(w(t))$ is strictly decreasing in t on the interior points unless $s_j(t) = 0$ for all j.

Now consider the other case where $w(t)$ is a boundary point. Assume that the bottleneck sets of w during $(t-\epsilon,\, t)$ and $(t,\, t+\epsilon)$ are B_1 and B_2, respectively, for small $\epsilon > 0$.

$$\begin{aligned}&\frac{V(w(t+\epsilon)) - V(w(t-\epsilon))}{2\epsilon}\\ &\qquad= \frac{V(w(t+\epsilon)) - V(w(t))}{2\epsilon} + \frac{V(w(t)) - V(w(t-\epsilon))}{2\epsilon}.\end{aligned}$$

Even though we define the Jacobian matrix J_x only on interior points, this can be extended to boundary points. On boundary points, we can define J_x as a function of direction d. From Corollary 3 in Appendix B, the right-hand directional derivative is well defined on the boundary points for an arbitrary direction d. Hence, depending on to which bottleneck sets $w(t-\epsilon)$ and $w(t+\epsilon)$ belong, the corresponding J_x can be used. Define $\dot{V}_B$ be a right-hand derivative of V defined on the bottleneck set B. Then, the above expression can be written as $(1/2)(\dot{V}_{B_1}(t) + \dot{V}_{B_2}(t))$ as ϵ approaches 0. Since both $\dot{V}_{B_1}(t)$ and $\dot{V}_{B_2}(t)$ are negative, the above expression is also negative.

We assumed that there exists an ϵ such that $w(s)\, s \in (t,\, t+\epsilon)$ lies in the same bottleneck. If a trajectory of $w(t)$ oscillates between boundaries infinitely many times, it may not be possible to find such an ϵ. However, from the continuity of $\dot{w}$, it can be shown such an infinite oscillation cannot happen.

Therefore, we have shown that V is a decreasing function of time t unless $s_j(t) = 0$. The theorem follows from [20]. ■

In [18], Kelly *et al.* proposed rate-based proportionally fair algorithm. Our algorithm is similar to Kelly's in that both achieve proportional fairness. However, it differs from Kelly's algorithm in that it is a window-based control and it does not need feedback from the routers. Kelly's algorithm changes the rate as follows:

$$\frac{d}{dt}x_i(t) = \kappa(p_i - x_i(t)A_i.\mu_j(t))$$

where $\mu_j(t) = ((A^T x)_j - C_j - \epsilon)/\epsilon^2$. The source i gets explicit feedback $\mu_j(t)$, residual capacity, from the links and changes its rate accordingly. The increase is linear and the decrease is multiplicative. Each $\mu_j(t)$ plays the role of a Lagrange multipliers of the problem P as $\epsilon \to 0$.

Our algorithm, however, controls the window size instead of the rate explicitly. Note that

$$\frac{d}{dt}w_i(t) = \kappa\left(\frac{p_i}{w_i} + \frac{d_i}{\overline{d}_i} - 1\right)\frac{d_i}{\overline{d}_i}, \quad \text{where } \overline{d}_i = d_i + \sum_{j\in A_{\cdot i}} q_j.$$

Here, the measured delay $\overline{d}_i$, which is the summation of q_j plus d_i, plays the role of implicit feedback. Note also that q_j in our algorithm is comparable to μ_j in Kelly's. They are both Lagrange multipliers of (P). However, we do not linearly increase and multiplicatively decrease the window. When the network is not congested, $q = 0$, $\dot{w} = \kappa(p_i/w_i)$. The increasing rate is a decreasing function of w.

C. (p, α)-Proportionally Fair Algorithm

In this subsection, we consider an algorithm that converges to the (p, α)-proportionally fair rate vector for $\alpha > 1$. We know

that if $s_i^\alpha = w_i - x_i d_i - (p_i/x_i^{\alpha-1}) = 0$ for all i, then the rate vector is (p, α)-proportionally fair. We call $p_i/x_i^{\alpha-1}$ the "target queue length," since $w_i - x_i d_i$ is the estimated queue length in the network. Note that the target queue length goes to infinity when the rate is very small. When $\alpha = 1$, the target queue length is constant regardless of the rate. On the other hand, when $\alpha > 1$, the target queue length is a function of x, which is varying and is a decreasing function of the rate. Hence, when the flow rate is large, the algorithm tries to maintain smaller queues and vice versa.

One unfavorable property of the target queue length function $p_i/x_i^{\alpha-1}$ is that when $x_i < 1$, this function becomes very large and the target queue length fluctuates and makes the control unstable. Consequently, we consider $p_i/(x_i+1)^{\alpha-1}$ instead of $p_i/x_i^{\alpha-1}$, since the variation of the former is smaller than that of the latter.

The objective function h_α such that the solution of (P) corresponds to $\overline{s}^\alpha = w_i - x_i d_i - (p_i/(x_i + 1^{\alpha-1})) = 0$ is

$$h_\alpha(x) = \begin{cases} \log x & \text{if } \alpha = 1 \\ \log\left(\dfrac{x}{x+1}\right) & \text{if } \alpha = 2 \\ \log\left(\dfrac{x}{x+1}\right) + \displaystyle\sum_{i=1}^{\alpha-2} \frac{1}{i(x+1)^i} & \text{if } \alpha = 3, 4, \ldots. \end{cases}$$

Note that $h'_\alpha = 1/(x(x+1)^{\alpha-1})$ and $\lim_{x\to\infty} h_\alpha = 0$. These observations show that $h_a p$ is increasing concave and nonnegative, and by the Claim 3 (Appendix B), the solution of (P) with objective function h_α converges to max–min rate vector.

Consider the system of differential equations

$$\frac{d}{dt} w_i = -\kappa x_i s_i u_i \tag{24}$$

where

$$s_i = w_i - x_i d_i - \frac{p_i}{(x_i+1)^{\alpha-1}} \tag{25}$$

$$u_i = d_i - (\alpha - 1)\frac{p_i}{(x_i+1)^\alpha}. \tag{26}$$

Theorem 6: If $p_i < d_i/(\alpha-1)$ for all i, the function $V(w) = (1/2)\sum_i s_i^2$ is a *Lyapunov function* for the system of equations (24)–(26). The unique value w minimizing $V(w)$ is a stable point of the system, to which all trajectories converge.

Proof: Note that

$$\begin{aligned} \frac{d}{dt} V(w(t)) &= \sum_j \frac{\partial V}{\partial w_j} \cdot \frac{dw_j(t)}{dt} \\ &= \sum_j \sum_i s_i \frac{ds_i}{dw_j} \dot{w}_j \\ &= s^T J_s \dot{w} \end{aligned}$$

where $J_s = ((ds_i/dw_j) i = 1, \ldots, N, j = 1, \ldots, N)$ is the Jacobian of s with respect to w. Equation (24) can be rewritten in a matrix form as $\dot{w} = -\kappa X U s$ where $U = \mathrm{diag}(u_i, i = 1, \ldots, N)$. If we show that $J_s X U$ is positive definite, then $V(w(t))$ is strictly decreasing with t, unless $s = 0$, the unique (p, α)-proportionally fair point.

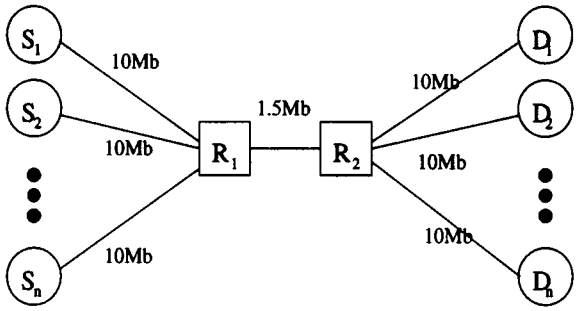


Fig. 3. Network topology.

$$\frac{\partial s_i}{\partial w_j} = \delta ij - d_i \frac{\partial x_i}{\partial w_j} + (\alpha - 1) p_i (x_i + 1)^{-\alpha} \frac{\partial x_i}{\partial w_j}$$

or

$$\begin{aligned} J_s &= I - D J_x + (\alpha - 1) P (X + I)^{-\alpha} J_x && (28) \\ &= I - U J_x && (29) \\ &= (I - U\overline{D}^{-1}) + U\overline{D}^{-1} X A J_q && (30) \end{aligned}$$

where $U = \mathrm{diag}(u_i, i = 1, \ldots, N)$ and $J_x = ((\partial x_i/\partial w_j), i \in \mathcal{N}, j \in \mathcal{L})$. Hence

$$\begin{aligned} J_s X U &= X U - (D - (\alpha - 1) P X) J_x X U \\ &= (I - D\overline{D}^{-1}) X U + (\alpha - 1) P (X + I)^{-\alpha} \overline{D}^{-1} X U \\ &\quad + U\overline{D}^{-1} X A J_q X U. \end{aligned}$$

Since $\overline{D}^{-1} X A J_q X$ is positive semidefinite, $(I - D\overline{D}^{-1})$ is a diagonal matrix with nonnegative entries, and $P(X + I)^{-\alpha}$ is positive definite, $J_s X U$ is positive definite.

By applying the same arguments of the previous proof for boundary points, we complete the proof. ■

The algorithm (24)–(26) is more of theoretical interest than practical. The difficulties such as variable queue size prevent this control from being implemented. However, it is noteworthy that max–min can be approximated in an end-to-end manner.

V. Simulation Results

In this section, we describe simulation results of our window-based algorithms to validate their performance. We used the *ns* simulator developed at the Lawrence Berkeley National Laboratory [7].

A. Network Topology

Fig. 3 shows the topology of the network which will be used to demonstrate the fairness and service differentiation of our algorithm of Section IV-B. In Fig. 3, the squares denote finite-buffer switches and the circles denote the end-hosts. Connection i transmits packets from S_i to D_i, and each connection passes bottleneck links between the routers R_1 and R_2. The links are labeled with their capacities. The propagation delays between S_i and R_1 and that between R_1 and R_2 are 1 msec and between R_2 and D_i is $8 + (i-1) \cdot 5$ msec. We choose different

TABLE I
THROUGHPUT OF 5 CONNECTION

connection	Fair	TCP-Reno
1	3645	6723
2	3479	2520
3	3510	2159
4	3668	2871
5	4373	4030

TABLE II
THROUGHPUT DIFFERENTIATION WITH DIFFERENT TARGETS

target(p_i)	throughput	throughput ratio	target ratio
2	420	1	1
6	1117	2.66	3
10	1863	4.43	5
14	2628	6.26	7
18	3353	7.98	9

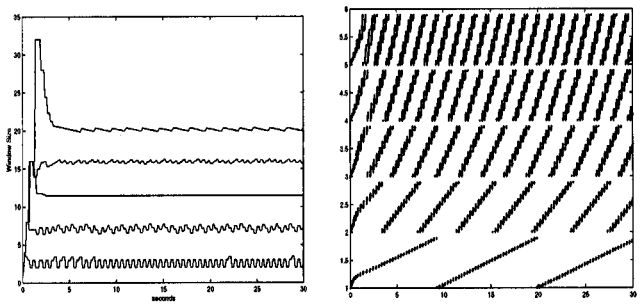


Fig. 4. Plot of window and packets with $p_i = 2$, 6, 10, 14, and 18.

propagation delays to show that our algorithm does not suffer from round-trip delay bias.

B. Fairness

We ran simulations for five connections with p_i of two packets for 100 s. The packet size is 1 KB. We measure the throughput achieved by those five connections. Table I shows the results. The second and third columns show the throughputs of connections when $(p, 1)$-fair algorithms and TCP-Renos are used respectively. It can be observed that our algorithm achieves fairer throughput without regard to the delay while the throughput of TCP-Reno varies more.

We use a different value of p_i for each connection. Table II shows the throughput of each connections with $p_i = 2$, 6, 10, 14, and 18 packets for 60 s.

The buffer size is of the router R_1 is 100 packets and the propagation delay of each connection is now set to 3 ms.

The left-hand side of Fig. 4 plots the window size over time of the five connections. The bottom line corresponds to the connection with target equal to 2 and the top line to that with target equal to 18. Note that a bigger target results in a bigger window size.

The right-hand side of Fig. 4 and Table II show that the throughput of connection i is almost proportional to p_i. The second column in Table II is the number of packets acknowledged for the five connections during the last 50 s. We omit the first 10 s to remove the effect of the slow start. Comparing the third column with the fourth shows that throughput is almost proportional to the target. This simulations shows that by controlling the target backlog (p_i), we can control the throughput of each connection.

VI. CONCLUSION

In this paper, we have addressed the fundamental question of the existence of fair end-to-end window-based congestion control. We have shown the existence of window-based fair end-to-end congestion control using a multiclass closed fluid model. We showed that the flow rate vector is a well-defined function of the window-size vector and characterized this function. We generalized the proportional fairness and related the fairness to the optimization problem. Our definition of fairness addresses the compromise between user fairness and resource utilization. With the help of an optimization problem, we have related window sizes and the fair rates. We have developed an algorithm which converges to the fair rates and proved its convergence using a Lyapunov function.

Our algorithm uses the propagation delay d_i, measured delay $\overline{d}_i$, and window size w_i. Unfortunately, the end user cannot know the exact value of propagation delay. Furthermore, the value of propagation delay could change in the case of rerouting in packet-switched networks. TCP-Vegas uses the minimum of delays observed so far as an estimated propagation delay. TCP-Vegas fails to adapt to the route change when the changed route is longer than original route. Refer to La *et al.* [26], [19] for this problem. The impact of feedback delays on the convergence will be also be topic of further studies.

APPENDIX A
UNIQUENESS OF THE RATE VECTOR

Theorem 2: Given (w, A, d, c), the flow rate vector x satisfying (1)–(4) is unique. We use two lemmas in the proof of the Theorem 2. The first one is a partial result of Rosen [27] and the second is Farkas' lemma (see, e.g., [24]).

Lemma 5: Let $F = (f_1, \ldots, f_n)$ be a vector of real-valued functions defined on $\mathcal{R}^n$. If the Jacobian matrix $\Delta F(x)$ exists and is either positive definite for all $x \in \mathcal{R}^n$ or negative definite for all $x \in \mathcal{R}^n$, then there is at most one x such that $F(x) = 0$ holds.

Proof: Assume there are two distinct points x^1 and x^2 such that $F(x^i) = 0$ for $i = 1, 2$. Let $x(\theta) = x^1 + \theta(x^2 - x^1)$ for $\theta \in [0, 1]$. Since ΔF is the Jacobian of F, we have

$$\frac{dF(x(\theta))}{d\theta} = \Delta F(x(\theta))\frac{dx(\theta)}{d\theta} = \Delta F(x(\theta))(x^2 - x^1).$$

Hence

$$F(x^2) - F(x^1) = \int_0^1 \Delta F(x(\theta))(x^2 - x^1)\, d\theta.$$

Multiplying both sides by $(x^2 - x^1)^T$ gives

$$(x^2 - x^1)^T(F(x^2) - F(x^1)) = \int_0^1 (x^2 - x^1)^T \Delta F(x(\theta))(x^2 - x^1)\, d\theta.$$

The left hand side of the above equation is 0 and the right hand side is either positive or negative depending on whether $\Delta F(x(\theta))$ is always positive definite or always negative definite. This contradiction completes the proof of the lemma. ■

Lemma 6 [Farkas]: $Ax = b$, $x \geq 0$ has no solution if and only if $yA \geq 0$, $yb < 0$ has a solution.

We are now ready to prove Theorem 2.

Proof of Theorem 2: The proof is composed of two parts. In the first part, we show that given (w, A, d, c), the set of bottleneck links B is uniquely determined. In the second part, we show that, given (w, A, d, c) and B, the vector of flow rates x is unique.

Claim 1: Given (w, A, d, c), the set of bottleneck links B defined by $B = \{j|(A^T x)_j = c_j\}$ is the same for all x that satisfies (1)–(4).

Assume that there exist two different sets of bottleneck links $B_1 \neq B_2$ that correspond to two distinct solutions (x^1, q^1) and (x^2, q^2) of (1)–(4), respectively. By (2), the queue size at a nonbottleneck link is 0. For $k = 1, 2$, designate by $\overline{q}^k$ the subvector of q^k with nonzero components. Let also A_k be the submatrix of A that consists of the columns of A that correspond to the nonzero components of q^k. With this notation we can write

$$Aq^k = A_k \overline{q}^k, \quad \text{for } k = 1, 2. \tag{31}$$

Plugging (31) into (3) and multiplying by $(X^k)^{-1}$, we find

$$A_k \overline{q}^k + d = (X^k)^{-1} w, \quad \text{for } k = 1, 2. \tag{32}$$

We partition the users into two sets $N^+ = \{i|x_i^1 \geq x_i^2\}$ and $N^- = \{1, \ldots, N\} \backslash N^+$ and rewrite (32) as

$$\begin{bmatrix} A_1^+ & -A_2^+ \\ -A_1^- & A_2^- \end{bmatrix} \begin{bmatrix} \overline{q}^1 \\ \overline{q}^2 \end{bmatrix} = \begin{bmatrix} (X^{1+} - X^{2+})^{-1} w^+ \\ (X^{2-} - X^{1-})^{-1} w^- \end{bmatrix} \tag{33}$$

by subtracting the identity (32) for $k = 2$ from the same identity for $k = 1$. The superscript $+$ and $-$ corresponds to the sets N^+ and N^-. Note that the right side of the (33) is less than or equal to 0. From Lemma 6, if there is a row vector $y = (y^+, y^-)$ such that

$$y^+ A_1^+ - y^- A_1^- \geq 0 \tag{34}$$

$$y^+ A_2^+ - y^- A_2^- \leq 0 \tag{35}$$

$$y^+ (X^{1+} - X^{2+}) w^+ + y^- (X^{2-} - X^{1-}) w^- < 0 \tag{36}$$

no $(\overline{q}^1, \overline{q}^2)$ satisfying (33) exists. We will show that $y = (y^+, y^-)$ with $y^+ = (x^{1+} - x^{2+})^T$ and $y^- = (x^{2-} - x^{1-})^T$ is such a vector.

Plug (y^+, y^-) defined above into (34), (35).

$$\begin{aligned} x^{1+} A_1^+ + x^{1-} A_1^- - x^{2+} A_1^+ - x^{2-} A_1^- &= x^1 A_1 - x^2 A_1 \\ &= c_{B_1} - x^2 A_1 \geq 0 \\ x^{1+} A_2^+ + x^{1-} A_2^- + x^{2+} A_2^+ + x^{2-} A_2^- &= x^1 A_2 - x^2 A_2 \\ &= x^1 A_2 - c_{B_2} \leq 0. \end{aligned}$$

We drop the superscript ^{T} in x^T for simplicity. The inequalities hold by inequality (1), hence (34) and (35) hold. For (36), note that the right-hand side of (33) is nonnegative and y is nonnegative. Hence (36) holds with a possibility of equality. The strict inequality follows from the fact $x^1 \neq x^2$. This completes the proof of the claim.

Claim 2: Given (w, A, d, c) and the corresponding set of bottleneck links B, the flow rate vector x that solves (1)–(4) is unique.

For simplicity of notation, we do not consider nonbottleneck links. That is, we rewrite the equations where every one of the M links is a bottleneck. If $\text{rank}(A) = N$, then the equations $A^T x = c$ determine x uniquely. Now we consider the case when $\text{rank}(A) = k < N$. By renumbering the connections and the network nodes, we can write A as

$$A = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

where E is a $k \times k$ invertible matrix and G is an $(N - k) \times k$ matrix. We claim that

$$H = GE^{-1}F. \tag{37}$$

To see why the above identity must hold, note that the rightmost $N - k$ columns of A are linear combinations of the leftmost k columns. That is, there is some $k \times (N - k)$ matrix M such that

$$\begin{bmatrix} F \\ H \end{bmatrix} = \begin{bmatrix} E \\ G \end{bmatrix} M.$$

Consequently, $F = EM$ and $H = GM$. The first identity implies $M = E^{-1}F$ and the second then yields $H = GE^{-1}F$, as claimed.

Let x_E and x_G be the vectors of flow rates corresponding to E and G, respectively. From $A^T x = c$ we find

$$x_E = E^{T-1} c_E - E^{T-1} G^T x_G \tag{38}$$

where c_E is a sub-vector of c corresponds to E from (1) and (2). [In (38), the notation E^{T-1} designates $(E^{-1})^T = (E^T)^{-1}$.] Let $b = X^{-1} w - d$ or $b_i = (w_i / x_i) - d_i$ for all i. Combining this notation with (3), we find

$$Aq = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} q_E \\ q_F \end{bmatrix} = b = \begin{bmatrix} b_E \\ b_G \end{bmatrix} \tag{39}$$

where $b^T = (b_E^T, b_G^T) = ((b_1, \ldots, b_k), (b_{k+1}, \ldots, b_N))$.

Multiplying the upper part of (39) by E^{-1}, we get an expression for q_E in terms of q_F: $q_E = E^{-1} b_E - E^{-1} F q_F$. Plugging this expression into $Gq_E + Hq_F = b_G$, we find $GE^{-1} b_E + (H - GE^{-1}F) q_F = b_G$ which reduces to

$$GE^{-1} b_E = b_G$$

by (37). Let $\tilde{G} := GE^{-1}$ and $F(x_G) := \tilde{G} b_E - b_G$. Note that F is a function of x_G, since b_E and b_G are also functions of x_G by (38) and the definition of b. We use Lemma 5 to show that there is a unique x_G so that $F(x_G) = 0$.

For $i = 1, \ldots, N - k$ and $j = k + 1, \ldots, N$, we compute the partial derivative of F_i with respect to x_j as follows. Note that

$$F_i(x_G) = \tilde{G}_i . b_E - b_{k+1+i} \tag{40}$$

where $\tilde{G}_{i\cdot}$ is row i of $\tilde{G}$. Now, for $m = 1, \ldots, k$

$$\frac{\partial (b_E)_m}{\partial x_j} = -\frac{w_m}{x_m^2}\frac{\partial x_m}{\partial x_j}.$$

Using (38), we see that

$$\frac{\partial x_m}{\partial x_j} = -\tilde{G}_{mj}.$$

Combining the previous two identities, we get

$$\frac{\partial (b_E)_m}{\partial x_j} = \frac{w_m}{x_m^2}\tilde{G}_{mj}.$$

Consequently

$$\frac{\partial}{\partial x_j} b_E = \operatorname{diag}\left(\frac{w_i}{x_i^2}\right)\tilde{G}_{j\cdot}^T.$$

Using this result and (40), we find

$$\frac{\partial F_i(x_G)}{\partial x_j} = \tilde{G}_{i\cdot}\operatorname{diag}\left(\frac{w_i}{x_i^2}\right)\tilde{G}_{j\cdot}^T + \frac{w_{k+i}}{{x_j}^2}\delta_{ij}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Hence the Jacobian matrix $\Delta F(x_G)$ of F is

$$\tilde{G}\begin{bmatrix} \frac{w_1}{x_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{w_k}{x_k^2} \end{bmatrix}\tilde{G}^T + \begin{bmatrix} \frac{w_{k+1}}{x_{k+1}^2} & & 0 \\ & \ddots & \\ 0 & & \frac{w_N}{x_N^2} \end{bmatrix}$$

which is positive definite. From Lemma 5, there is a unique x_G so that $F(x_G) = 0$ and by (38), x is unique. ■

APPENDIX B
SOME PROPERTIES OF THE MAPPING

Claim 3: F: $W \to X$ is a continuous function of w.

Proof: Consider an optimization problem $\max_x f(w, x) = \sum_i w_i \log x_i - d_i x_i$ subject to (1) and (4). Then the Kuhn–Tucker conditions correspond to (1)–(4).[3] Hence, $F(w)$ is the optimal solution of the problem. Take a sequence w_n such that $w_n \to w$ and let $x_n = F(w_n)$. There exists a subsequence x_{n_k} that converges to, say $\overline{x}$, by the compactness of the constraint set. By the optimality of x_{n_k}

$$f(w_{n_k}, x_{n_k}) \geq f(w_{n_k}, x), \quad \text{for all } x.$$

Upon taking limits and invoking the continuity of f

$$f(w, \overline{x}) \geq f(w, x), \quad \text{for all } x.$$

This proves $\overline{x} = F(w)$ and therefore, $F(x)$ is continuous. ■

Claim 4: F is differentiable except at the boundary points.

Proof: Define $g_B(w, q) = A_B^T x(q_B) - c_B$, where $x(q_B) = w_i/(d_i + A_{i\cdot}q_B)$. If A_B has a full rank, i.e., $\operatorname{rank}(A_B) = |B|$, then by the implicit function theorem, $q_B(w)$ is differentiable, hence $x(w)$ is differentiable. Now assume $\operatorname{rank}(A_B) = r < |B|$. Let $\overline{B} \subset B$ such that $|\overline{B}| = r$. Then applying same implicit function theorem to $g_{\overline{B}}$ gives the differentiability of $x(q_{\overline{B}})$. To make the proof complete, observe that $x(q_{\overline{B}})$ is the unique solution of g_B. Since we delete dependent rows from B, if x is a solution of $A_{\overline{B}}^T x = c_{\overline{B}}$, it is also solution of $A_B^T x = c_B$. Since the solution x is unique, $x(q_{\overline{B}})$ is the solution of g_B. This completes the proof. ■

Corollary 3: Let $D_u^+ F = \lim_{\epsilon \downarrow 0}(F(w + \epsilon u) - F(w))/\epsilon$. Then $D_u^+ F$ exists for all w.

Proof: If w is an interior point, the conclusion is obvious. If w is a boundary point for any direction u, there exists $\epsilon > 0$ such that $\overline{w} \in (w, w + \epsilon u]$ has the same bottleneck $\overline{B}$. Then restricting domain of F to $(w, w + \epsilon u]$ and applying the same argument as in the proof of claim gives the result. ■

Lemma 4: The Jacobian $J_x = [(\partial x_i/\partial w_j), i, j \in \mathcal{N}]$ of $x(w)$ with respect to w is given by the following expression on the interior point:

$$J_x = \overline{D}^{-1}\left(I - XA_B(A_B^T X \overline{D}^{-1} A_B)^{-1} A_B^T \overline{D}^{-1}\right) \quad (41)$$

where

$$\overline{D} = \operatorname{diag}(d_i + A_{i\cdot}q, i \in \mathcal{N})$$

and

$$X = \operatorname{diag}(x_i, i \in \mathcal{N}). \quad (42)$$

Proof: Our starting point is (3) which reads

$$x_i\left[(A_B q_B)_i + d_i\right] = w_i, \; i \in B \quad (43)$$

where q_B is the subvector of q that corresponds to the bottleneck links. This equation contains the dependencies of x_i on w_j. Accordingly, we see that to compute J_x we need only consider the bottleneck links. We drop the subscript B from A_B and q_B in the rest of the proof. Let also c_B be the subvector of c that corresponds to the bottleneck links and we drop the subscript B. Without loss of generality, we can assume $\operatorname{rank}(A_B) = |B|$, since otherwise, we can reduce B to have a full rank, and from the proof of Theorem 2, the reduced system has the same solution as the original system.

With this notation, we have

$$x_i[A_{i\cdot}q + d_i] = w_i \quad (44)$$

$$A^T x = c. \quad (45)$$

Taking the partial derivative of (44) with respect to w_j we find

$$(J_x)_{ij}(A_{i\cdot}q + d_i) + x_i(A_{i\cdot}(J_q)_{\cdot j}) = \delta_{ij}.$$

We can write this identity as follows:

$$\overline{d}_i(J_x)_{ij} + x_i(A_{i\cdot}(J_q)_{\cdot j}) = \delta_{ij}.$$

In matrix notation, these identities read

$$\overline{D}J_x + XAJ_q = I. \quad (46)$$

Multiplying this identity to the left by $A^T\overline{D}^{-1}$, we find

$$A^T J_x + (A^T\overline{D}^{-1}XA)J_q = A^T\overline{D}^{-1}. \quad (47)$$

[3]This observation is credited to S. Low. It also can simplify the proof of Theorem 1 and 2.

Now, (45) implies that $A^T J_x = 0$. Hence

$$J_q = (A^T \overline{D}^{-1} X A)^{-1} A^T \overline{D}^{-1} \tag{48}$$

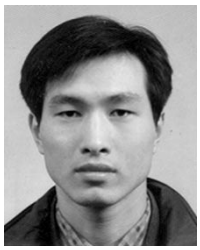
Plugging (48) into (46) gives (41). ■

ACKNOWLEDGMENT

The authors would like to thank J.-Y. Le Boudec for his helpful criticism and give credit to him for the proof of Lemma 3. The authors also express their gratitude to R. La for helping to complete the work. Finally, they thank S. Floyd, S. Low, and the anonymous reviewers for their comments to improve the paper.

REFERENCES

[1] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[2] A. Charny, "An algorithm for rate allocation in a packet-switching network with feedback," Master's thesis, Mass. Inst. Technol., Cambridge, MA, 1994.
[3] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Comput. Networks ISDN Syst.*, vol. 17, pp. 1–14, 1989.
[4] S. Floyd and V. Jacobson, "Connection with multiple congested gateways in packet-switched networks, Part 1: One-way traffic," *ACM Comput. Commun. Rev.*, vol. 21, no. 5, pp. 30–47, Aug. 1991.
[5] ——, "On traffic phase effects in packet switched gateways," *Internetworking: Res. and Experience*, vol. 3, no. 3, pp. 115–156, Sept. 1993.
[6] ——, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Networking*, vol. 1, pp. 397–413, Aug. 1993.
[7] S. McCanne, S. Floyd, and K. Fall. (1997) *ns*–LBNL Network Simulator. Lawrence Berkeley National Laboratory Networking Group, Berkeley, CA. [Online]. Available: http://www-nrg.ee.lbl.gov/ns/
[8] E. Hahne, "Round-robin scheduling for max–min fairness in data network," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 1024–39, Sept. 1991.
[9] T. Henderson, E. Sahouria, S. McCanne, and R. Katz, "Improving fairness of TCP congestion avoidance," in *Globecom'98*, Sydney, Australia, pp. 539–544.
[10] P. Hurley, J.-Y. Le Boudec, and P. Thiran, "A note on the fairness of additive increase and multiplicative decrease," in *Proc. ITC-16*, Edinburgh, Scotland, June 1999, pp. 467–478.
[11] V. Istratescu, *Fixed Point Theory*. Dordrecht, Holland: Reidel, 1981.
[12] V. Jacobson, "Congestion avoidance and control," *Comput. Commun.n Rev.*, vol. 18, no. 4, pp. 314–29, Aug. 1988.
[13] J. M. Jaffe, "Bottleneck flow control," *IEEE Trans. Commun.*, vol. COM-29, July 1981.
[14] ——, "Flow control power is nondecentralizable," *IEEE Trans. Commun.*, vol. COM-29, pp. 1301–1306, Sept. 1981.
[15] R. Jain, "Myths about congestion management in high-speed networks," *Internetworking: Res. and Experience*, vol. 3, pp. 101–113, 1992.
[16] ——, "Congestion control and traffic management in ATM networks: Recent advances and a survey," *Comput. Networks ISDN Syst.*, vol. 28, no. 13, pp. 1723–38, Oct. 1996.
[17] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, pp. 33–37, Jan./Feb. 1997.
[18] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow price proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.
[19] R. J. La, J. Walrand, and V. Anantharam. (1998, July) Issues in TCP vegas. [Online]. Available: http://www.path.berkeley.edu/~hyongla
[20] D. Luenberger, *Introduction to Dynamic System*. New York, NY: Wiley, 1979.
[21] C. Boutremans, M. Vojnovic, and J. Y. Le Boudec, "Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times," in *IEEE Infocom'00*, Tel Aviv, Israel, Mar. 2000, pp. 1303–1312.
[22] A. Mankin, "Random drop congestion control," in *Proc. SIGCOMM'90*, Philadelphia, PA, Sept. 1990, pp. 1–7.
[23] L. Massoulie and J. Roberts, "Bandwidth sharing: Objectives and algorithms," in *IEEE Infocom'99*, vol. 3, New York, NY, Mar. 1999, pp. 1395–1403.
[24] M. Avriel, *Nonlinear Programming*. Englewood Cliffs, NJ: Prentice-Hall, 1976.
[25] A. Mayer, Y. Ofek, and M. Yung, "Approximating max–min fair rates via distributed local scheduling with partial information," in *IEEE Infocom'96*, San Francisco, CA, USA, Mar. 1996, pp. 926–936.
[26] J. Mo, R. J. La, J. Walrand, and V. Anantharam, "Analysis and comparison of Reno and Vegas," in *IEEE Infocom'99*, vol. 3, New York, NY, Mar. 1999, [Online]. Available: http://www.path.berkeley.edu/~jhmo, pp. 1556–1563.
[27] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
[28] S. Shenker, "A theoretical analysis of feedback flow control," in *ACM SIGCOM'90*, Philadelphia, PA, Sept. 24–27, 1990, pp. 156–165.



Jeonghoon Mo received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea. He received the M.S. and Ph.D. degrees from the University of California, Berkeley.

Currently, he is a Senior Technical Staff Member with AT&T Labs, Middletown, NJ. His research interests include communication networks, queueing theory, Internet protocols, and quality-of-service issues.

Jean Walrand (S'71–M'80–SM'90–F'93) is a Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. His research interests include stochastic processes, queueing theory, communication networks, and control systems. He is the author of *An Introduction to Queueing Networks* (Prentice Hall, 1988), *Communication Networks: A First Course* (McGraw-Hill, 1998), and co-author of *High-Performance Communication Networks* (Morgan Kaufmann, 2000).

Dr. Walrand is a recipient of the Lanchester Prize from the Operations Research Society of America and of the S. Rice Prize from the Communication Society of IEEE.