FAIR END-TO-END WINDOW-BASED CONGESTION CONTROL

by

Jeonghoon Mo and Jean Walrand

Memorandum No. UCB/ERL M98/46

15 July 1998

cover

FAIR END-TO-END WINDOW-BASED CONGESTION CONTROL

by

Jeonghoon Mo and Jean Walrand

Memorandum No. UCB/ERL M98/46

15 July 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Fair End-to-End Window-based Congestion Control

Jeonghoon Mo and Jean Walrand
University of California at Berkeley
{jhmo,wlr}@eecs.berkeley.edu

July 15, 1998

Abstract

In this paper, we demonstrate the existence of fair end-to-end window-based congestion control protocols for packet-switched networks with FCFS 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 construction of the protocols.

1 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 nor 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 ([15]), 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 [10, 12, 23]. A bottleneck is a node where packets are backlogged so that its transmission rate limits the rate of the connections that go through it. The observed bias can be explained as follows. While a host does not detect congestion, it increases its window size by one unit per round-trip time of the connection. Accordingly, the window size of a connection with a short propagation delay increases faster than that of a connection with a longer propagation delay. Consequently, a long-delay connection looses out when competing with a short-delay connection.

Based on this observation, Floyd and Jacobson [11] proposed a "constant rate adjustment" algorithm. Handerson et al [13] simulated a variation of this scheme. They report that if the rate of increase of the window size is not excessive, then this scheme is not harmful to the other connections that use the original TCP scheme. Moreover, as expected, this scheme results in better performance for connections with longer propagation time. 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 [17] shows that power cannot be optimized in a distributed manner.

Chiu and Jain [4] 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 [15]. However, these implementations control the size of their window and not their transmission rate. Moreover, simple examples show that the result does not hold for networks with multiple bottleneck nodes.

Shenker [29] 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 [29] to justify the necessity of switch-based control for fairness [27, 3].

Recently, Kelly et al [20] exhibited an aggregate feedback algorithm that converges to a *proportionally fair* 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. Window-based algorithms are used to control errors. A rate-based control must be augmented with another retransmission protocol for error control. The window-based control algorithm integrates these two functions of error and congestion control. Although the delay is an additive congestion signal, this is less informative than Kelly's. Our protocol can be viewed as a refinement of TCP congestion control algorithms.

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.

2 MODEL

Window flow control is usually modeled as a closed queuing network [5, 6, 9]. For instance, in [5], the authors study the window flow control of *a single connection* with fixed propagation delay in a product form network. They derived the optimal window size and an adaptive window control scheme based on the analytical model.

In this paper we consider *a closed multiclass fluid network* with M links and N connections. We define that model next. The sender of connection i ($i = 1, \ldots, N$) 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 ($j = 1, \cdots, M$) has capacity, or transmission rate, c_j. We define the matrix $A = (A_{ij}, i = 1, \ldots, N, j = 1, \cdots, M)$ 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 acknowledgement by the same host. We assume that the processing times are negligible. A typical acknowledgement delay comprises d_i and some additional queuing delay in bottleneck routers. Let x_i be the flow rate of the i-th connection for $i = 1, \ldots, N$. For $j = 1, \cdots, M$, 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 (FCFS).

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

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

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

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

$$x \geq 0, \; q \geq 0 \tag{4}$$

where

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

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

$$q_j[(A^T x)_j - c_j] = 0, \text{ for } j = 1, \ldots, M.$$

The j-th 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 identities (3), which can be written as

$$x_i[(Aq)_i + d_i] = w_i, i = 1, \ldots, N,$$

mean that the total number of packets w_i for each connection $i, i = 1, \ldots, N$, 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 identities (3) as follows:

$$x_i = \frac{w_i}{D_i} \text{ where } D_i = d_i + (Aq)_i \text{ for } i = 1, \ldots, N. \tag{5}$$

The identities (5) mean 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 queuing 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.

We are going to prove that the flow rates x are a well-defined function of the window sizes w. 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 $j = 1, \cdots, M$. For $q \in [0, U]^M$, let

$$x_i(q) := \frac{w_i}{d_i + (Aq)_i}, i = 1, \ldots, N, f(q) := c - A^T x(q), \text{ and } h_j(q) = -f_j^2(q), j = 1, \cdots, M.$$

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 $f_j(0, q^j) < 0$ 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 [14], 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

$$q_j^* = \arg\max_{q_j \in [0,U]} h_j(q_1^*, \ldots, q_{j-1}^*, q_j, q_{j+1}^*, \ldots, q_M^*) \text{ for } j = 1, \cdots, M. \tag{6}$$

Let q^* be such that (6) holds and let $x^* = x(q^*)$. We claim is 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 2 *Given (w, A, d, c), the flow rate $x = (x_1, \cdots, x_M)$ satisfying the equations (1)- (4) is unique.*

We use two lemmas in the proof of the theorem 2. The first one is a partial result of Rosen [28] and the second is Farkas' lemma (see e.g., [24]).

Lemma 1 *Let $F = (f_1, \cdots, f_n)$ 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

$$(x^2 - x^1)^T(F(x^2) - F(x^1)) = \int_0^1 (x^2 - x^1)^T \Delta F(x(\theta))(x^2 - x^1)d\theta. \tag{7}$$

The left side of the equation (7) 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 2 (Farka) *$Ax = b, x \geq 0$ has no solution if and only if $yA \geq 0, yb < 0$ has a solution.*

Proof: For a proof, see [24]. ∎

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 the equations (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 the equations (1)-(4), respectively. By the equation (2), the queue size at a non-bottleneck link is 0. For $k = 1, 2$, designate by $\bar{q}^k$ 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

$$Aq^k = A_k \bar{q}^k \text{ for } k = 1, 2. \tag{8}$$

Plugging (8) into (3) and multiplying $(X^k)^{-1}$,

$$A_k \bar{q}^k + d = (X^k)^{-1} w \text{ for } k = 1, 2 \tag{9}$$

We partition the users into two sets $N^+ = \{i|x_i^1 \geq x_i^2\}$ and $N^- = \{1, \cdots, N\} \backslash N^+$ and rewrite (9) as

$$\begin{bmatrix} A_1^+ & -A_2^+ \\ -A_1^- & A_2^- \end{bmatrix} \begin{bmatrix} \bar{q}^1 \\ \bar{q}^2 \end{bmatrix} = \begin{bmatrix} (X^{1+} - X^{2+})^{-1} w^+ \\ (X^{2-} - X^{1-})^{-1} w^- \end{bmatrix} \tag{10}$$

by subtracting the equation (9) for $k = 2$ from the same equation for $k = 1$. The superscript $+$ and $-$ corresponds to the sets N^+ and N^-. Note that the right side of the (10) is less than or equal to 0. From the lemma ??, if there is a row vector $y = (y^+, y^-)$, such that

$$y^+ A_1^+ - y^- A_1^- \geq 0 \tag{11}$$

$$y^+ A_2^+ - y^- A_2^- \leq 0 \tag{12}$$

$$y^+ (X^{1+} - X^{2+}) w^+ + y^- (X^{2-} - X^{1-}) w^- < 0, \tag{13}$$

no $(\bar{q}^1, \bar{q}^2)$ satisfying the equation (10) 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.

Plugg (y^+, y^-) defined above into (11)-(12).

$$x^{1+} A_1^+ + x^{1-} A_1^- - x^{2+} A_1^+ - x^{2-} A_1^- = x^1 A_1 - x^2 A_1 = c_{B_1} - x^2 A_1 \geq 0$$
$$x^{1+} A_2^+ + x^{1-} A_2^- + x^{2+} A_2^+ + x^{2-} A_2^- = x^1 A_2 - x^2 A_2 = x^1 A_2 - c_{B_2} \leq 0$$

We drop the superscript ^{T} in x^T for simplicity. The inequalities holds by equation (1), hence (11) and (12) holds. For (13), note that the right hand side of (10) is nonnegative and y is nonnegative. Hence (13) 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 a the corresponding set of bottleneck links B, the flow rate x that solves the equations (1)-(4) is unique*

For simplicity of notation, we do not consider non-bottleneck links. That is, werewrite the equations where every one of the M links is a bottleneck. If $rank(A) = N$, then the equations $A^T x = c$ determine x uniquely. Now we consider the case when $rank(A) = k < N$. 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 a $(N-k) \times k$ matrix. We claim that

$$H = GE^{-1}F. \tag{14}$$

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

$$x_E = E^{T-1}c_E - E^{T-1}G^T x_G \tag{15}$$

where c_E is a sub-vector of c corresponds to E from the equations (1) and (2). (In (15), the notation E^{T-1} designates $(E^{-1})^T = (E^T)^{-1}$.) Let $b = X^{-1}w - d$ or $b_i = \frac{w_i}{x_i} - d_i$ for $i = 1, N$. Combining this notation with equation (3), we find

$$Aq = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \begin{bmatrix} q_E \\ q_F \end{bmatrix} = b = \begin{bmatrix} b_E \\ b_G \end{bmatrix} \tag{16}$$

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

Multiplying E^{-1} to the upper part of matrix equation (16), we get an expression for q_E in terms of q_F: $q_E = E^{-1}b_E - E^{-1}Fq_F$. 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 (14). 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 function of x_G by equation (15) and the definition of b. We use Lemma 1 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

$$F_i(x_G) = \tilde{G}_{i\cdot} b_E - b_{k+1+i} \tag{17}$$

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 (15), 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,

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

Using this result and the expression (17), we find

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

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 the lemma 1, there is a unique x_G so that $F(x_G) = 0$ and by the equation (15), x is unique ∎

Alhough the rate vector is uniquely determined from the window sizes, the workload vector generally q 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 $\frac{w}{d} > c$, then the queues build up in the links. For this network, any vector (q_1, q_2) such that $q_1 + q_2 = \frac{w}{c} - d$ is a solution of the equations (1) - (3).

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

Corollary 1 *If $rank(A_B)$ 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 equation (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. ∎

The following lemma provides sufficient conditions for links not to be bottlenecks.

Lemma 3 *For any given window size vector w, $0-1$ matrix A, and diagonal matrix D,*

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

Proof: (a) Assume that $q_j > 0$. This implies $x_i < \frac{w_i}{d_i}$. 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 the equation (2), $q_j = 0$, which is contradiction. Hence $q_j = 0$.

(b) If $q = 0$, the window size vector $w = Xd$ from (3) where $D = diag(d_i, i = 1, \cdots, N)$. By (1), we prove if part. The only if part is obvious from part (a). ∎

The converse of part (a) 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 $A_{\cdot 2}^T D^{-1} w = 10 > 5$.The end-to-end protocol that we develop controls the window sizes.

Let $F : W \rightarrow X$ be the mapping from the window space W to a flow rate space X defined by (1)-(4). F is a continuous function but is not always differentiable as the next example shows. Consider the network and connections in figure 1(a). Two users are sharing one link and each uses another link. Figure 1(b) is a plot of x_1 along the horizontal dotted line 1 in figure 1(c). Figure 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 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.

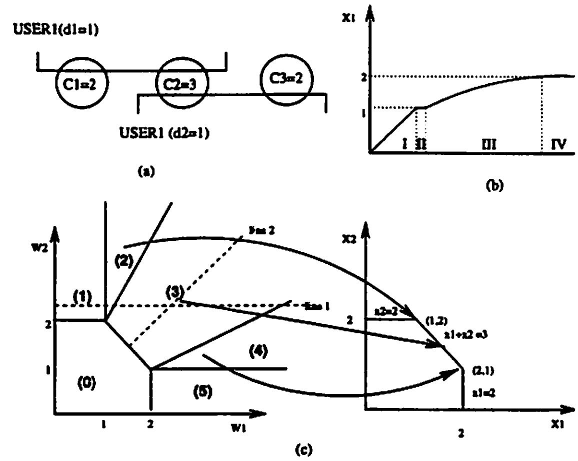

Figure 1: (a)network topology (b) flow rate x vs. window size (c) mapping between x and w

Figure 1(c) shows the mapping $x = F(w)$. If $w \in (0)$, there is no queue, 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 $XAq = XA_B q_B$, $F^{-1}(w)$ is a positive cone of XA_B with vertex Xd, as we now illustrate in figure 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 $F^{-1}(x)$.

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 $B(\bar{w})$ are same for all $\bar{w} \in$ neighborhood, $N_\epsilon(w)$, of w. Otherwise, w is said to be a boundary point.

Claim 3 *F is a continuous function of w.*

Proof: If w is an interior point, then by claim 4, F is continuous. Let w be an boundary points, i.e. $\bar{B}(w) = \{B(\bar{w}) | \bar{w} \in N_\epsilon(w)\}$ is not unique for small ϵ.

Let w_n be a sequence such that $w_n \rightarrow w$. Define $x_n = F(w_n)$. Take a subsequence n_k such that $B(w_{n_k})$ are same, say, $\hat{B}$ for all k. Then $x_{n_k} \rightarrow$ for some $\bar{x}$. It is enough to prove that $\bar{x} = x$, all the subsequences can be constructed as a combination of x_{n_k}. If $\hat{B} = B(w)$, that is the subsequence has same bottlenecks as w, $\bar{x} = x$. If $A_{\tilde{B}}$ has a full rank, $(q_{n_k})_j$ is a unique solution of $\max h_j(w_{n_k}, q_{n_k})$ where $h_j(w, q)$ same as in the proof of theorem 1.

$$h_j(w_{n_k}, (q_{n_k})_j, (q_{n_k})_{j-}) \geq h_j(w_{n_k}, y, (q_{n_k})_{j-}) \text{ for all } y$$

Upon taking limits and invoking continuity of h,

$$h_j(w^*, (q^*)_j, (q^*)_{j-}) \geq h_j(w^*, y, (q^*)_{j-}) \text{ for all } y$$

Hence, $\bar{x} = \frac{w_i}{d_i + (A\bar{q})_i} = x$. If $A_{\tilde{B}}$ does not have a full rank, select a submatrix $A_{\tilde{B}}$ such that it has a full rank and same range space. Define $(q_{n_k})_j = arg\max h_j$ if $j \in \tilde{B}, otherwise (q_{n_k})_j = 0$ Note that q_{n_k} is unique. Applying same arguement gives desired results. ∎

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

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

Corollary 2 *Let $D_u^+ F = \lim_{\epsilon \downarrow 0} \frac{F(w+\epsilon u) - F(w)}{\epsilon}$. Then $D_u^+ F$ exists for all w.*

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

3 FAIRNESS

3.1 Fairness

Fairness has been defined in a number of different ways so far. One of the most common fariness definitions is *max-min* or *bottleneck optimality* criterion [16, 1, 8, 18, 3]. 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, 18, 3]. But *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.

Kelly [19] proposed *proportionally 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:

$$\sum_i \frac{x_i - x_i^*}{x_i^*} < 0. \tag{18}$$

In [20], Kelly et al. suggested simple linear increase and multiplicative decrease algorithm converges to *proportionally fair* point.

Recently, game theory has been applied to flow control. [26, 7, 30]. These authors model users as players competing for common shared resources. The concept of *Nash Equilibria* provides a framework for definining fairness and proper operating points for the network. In [7], the game is viewed as non-cooperative. In [26], it is modeled as a cooperative game in which the users act to achieve better common utilities.

Next, we generalize the concept of *proportional fairness.* Consider the following optimization problem:

(P)

$$\text{maximize} \quad g = \textstyle\sum_i p_i f(x_i) \tag{19}$$

$$\text{subject to} \quad A^T x \leq c \tag{20}$$

$$\text{over} \quad x \geq 0 \tag{21}$$

where f is an increasing strictly concave function and the p_i are positive numbers. Since the objective function (19) is strictly concave and the feasible region (20)-(21) 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

$$\nabla g^T - \mu^T A^T = 0 \tag{22}$$

$$\mu_j(c_j - A_{\cdot j}^T x^*) = 0 \text{ for } j = 1, \cdots, M \tag{23}$$

$$A^T x^* \leq c \tag{24}$$

$$x^* \geq 0, \mu \geq 0 \tag{25}$$

where $\nabla g^T = (p_1 f'(x_1), \cdots, p_n f'(x_n))$. When there is only one link and N connections, the optimal solution of (P) is $x_i = \frac{c}{N}$ for all i: All the connections have an equal share of the bottleneck capacity, irrespective of the increasing concave f. Indeed, (22) 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 situation is not same when there are multiple bottlenecks. Consider the following network with 2 different bottlenecks and 3 connections. The *max-min fair*

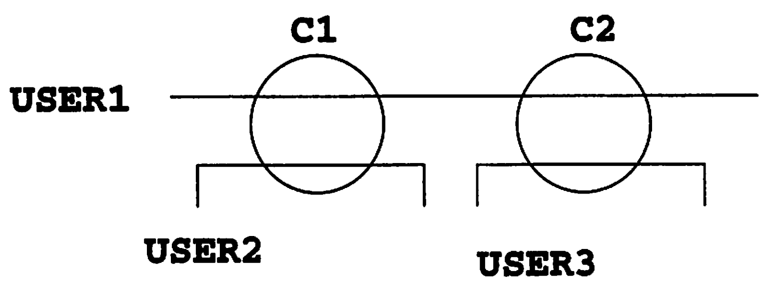

Figure 2: Network with multiple bottlenecks

rate vector of this network is $(\frac{c_1}{2}, \frac{c_1}{2}, c_2 - \frac{c_1}{2})$ if $c_1 < c_2$, while the *proportionally fair* rate vector is not same as *max-min 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 $x < y$.

It is a matter of controversy what is a fair rate allocation for the network in figure 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.

We generalize the concept of *proportional fairness* as follows.

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,*

$$\sum_i p_i \frac{x_i - x_i^*}{x_i^{*\alpha}} < 0. \tag{26}$$

Note that (26) reduces to (18) when $p = (1, \cdots, 1)^T$ and $\alpha = 1$.

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

Lemma 4 *Define the function f_α as follows:*

$$f_\alpha(x) := \begin{cases} (1-\alpha)^{-1}x^{1-\alpha} & \textit{if } 0 < \alpha < 1 \\ \log x & \textit{if } \alpha = 1 \\ -(1-\alpha)^{-1}x^{1-\alpha} & \textit{if } \alpha > 1 \end{cases}$$

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 (22) by $(x - x^*)$, we find $\nabla g^T(x - x^*) = \mu^T A^T(x - x^*)$. Summing the equations (23) over j, we obtain $\mu^T c = \mu^T A^T x^*$. Multiplying (20) 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 $\nabla g^T(x - x^*) = \sum_i p_i \frac{x - x^*}{x^{*\alpha}}$. Hence, $\sum_i p_i \frac{x - x^*}{x^{*\alpha}} \leq 0$. The strict inequality holds for all $x \neq x^*$, because of the uniqueness of x^*. We have shown that (26) holds and that x^* is (p,α)-proportionally fair.

To prove the converse, assume that x^* is (p,α)-proportionally fair. We show that it solves (P). First note that $g(x) = g(x^*) + \nabla g(x^*)^T(x - x^*) + o(x - x^*)$. Since $\nabla g(x^*)^T(x - x^*) = \sum_i p_i \frac{x - x^*}{x^{*\alpha}} < 0$ for all feasible x, it follows that 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 5 *If h is increasing concave negative function, the solution of (P) with $f_\alpha = -(-h)^\alpha$ approaches the* max-min fair *rate vector as $\alpha \to \infty$.*

Proof: Let x^α be the optimal solution of (P) with f_α and $\mathcal{X} = \{A^T x \leq c, x \geq 0\}$. 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 $x^\infty \in \mathcal{X}$ as $k \to \infty$.

By the optimality of x^{α_k}, for all feasible $y (\neq x^{\alpha_k})$,

$$\sum_i p_i f'_{\alpha_k}(x_i^{\alpha_k})(y_i - x_i^{\alpha_k}) < 0, \tag{27}$$

i.e.,

$$p_i f'_{\alpha_k}(x_i^{\alpha_k})(y_i - x_i^{\alpha_k}) < -\sum_{j \neq i} p_j f'_{\alpha_k}(x_j^{\alpha_k})(y_j - x_j^{\alpha_k}). \tag{28}$$

Dividing both sides of the equation (28) with $p_i(y_i - x_i^{\alpha_k})$,

$$f'_{\alpha_k}(x_i^{\alpha_k}) < \sum_{j \neq i} r_j f'_{\alpha_k}(x_j^{\alpha_k}) \leq \sum_{j \neq i, r_j > 0} r_j f'_{\alpha_k}(x_j^{\alpha_k}) \tag{29}$$

where $r_j = -\frac{p_j(y_j - x_j^{\alpha_k})}{p_i(y_i - x_i^{\alpha_k})} < \infty$.
Since $\left(\frac{h(x^{\alpha_k})}{h(x^{\alpha_k}+\epsilon)}\right) > 1$ and $\lim_{\alpha_k \to \infty} \left(\frac{h(x^{\alpha_k})}{h(x^{\alpha_k}+\epsilon)}\right) > 1$,

$$\frac{f'_\alpha(x^{\alpha_k})}{f'_\alpha(x^{\alpha_k}+\epsilon)} = \left(\frac{h(x^{\alpha_k})}{h(x^{\alpha_k}+\epsilon)}\right)^{\alpha_k - 1} \frac{h'(x^{\alpha_k})}{h'(x^{\alpha_k}+\epsilon)} \to \infty \text{ as } \alpha_k \to \infty \text{ for an arbitrary } \epsilon .$$

Taking limits on both sides of the equation (29), for this inequality to hold, there exists k such that $x_i^\infty \geq x_k^\infty$. Note that $r_k > 0$ implies if $y_i > x_i^\infty$ then $y_k < x_k^\infty$. By the definition of the max-min rate vector (x is max-min if for every feasible y, if $x_i > y_i$ there exists j such that $y_j < x_j \leq x_i$) x^∞ is the max-min rate vector. From the uniqueness of the max-min rate vector and that above arguments holds for an arbitrary subsequence, the proof is complete. ∎

Corollary 3 *The* (p, α)- proportionally fair *rate vector approaches the* max-min *fair rate vector as $\alpha \to \infty$.*

Proof: This result follows from Lemma 5 because f_α satisfies the conditions of the lemma. ∎

3.2 Window Size and Fairness

In this subsection, we study the relationship between window sizes and fairness.

TCP vegas [2] uses the estimated total backlog of a connection as a decision function. In our notation, the total backlog of connection i is $w_i - x_i d_i$. In TCP vegas, a host increases its window size if the estimated total backlog is smaller than a target value and decreases it otherwise.

We now establish the relationship between the total backlogs and fairness. Let $p_i > 0$ for $i = 1, \cdots, N$. Define

$$s_i = w_i - x_i d_i - p_i, \text{ for } i = 1, \cdots, N. \tag{30}$$

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 $i = 1, \cdots, N$. Moreover, the corresponding rate vector $x(w)$ defined by the equations (1)-(4) is a $(p, 1)$-proportionally fair rate vector.*

Proof: For any given $w \in [0, \infty)^N$, let x be the rate vector that corresponds to w, as defined by the equations (1)-(4). Fix $w \in W$. From equation (3), we get $XAq = w - Xd = p$ where $p = (p_1, \cdots, p_N)^T$. The last equality follows from the definition of W. Hence equation (22) is satisfied. From (1), (2), and (4), if we replace μ_j with q_j, the optimality conditions (23)-(25) of problem (P) hold for $f(x) = p_i \log x$ for $i = 1, \cdots, N$. By Lemma 4, 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, equation (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 $i = 1, \cdots, N$ and $\alpha > 1$. Define

$$s_i^\alpha = w_i - x_i d_i - \frac{p_i}{x_i^{\alpha-1}}, \text{ for } i = 1, \cdots, N. \tag{31}$$

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 equations (1)-(4) is a (p, α)-* proportionally fair *rate vector.*

Proof: Note that if $s_i^\alpha = 0$, then $w_i - x_i d_i = \frac{p_i}{x^{\alpha-1}} = x_i A_{i\cdot} q$. Consequently, $x^\alpha A q = p$, which is the optimality condition (22). The other conditions (24), (23), and (25) are satisfied as can be shown as in the previous proof. ∎

4 ALGORITHM

4.1 (p,1)-Proportionally Fair Algorithm

In this section, we construct an end-to-end control that converges to the *proportionally fair* point. Define

$$\bar{d}_i = d_i + A_{i\cdot} q, \text{ for } i = 1, \cdots, N.$$

That is, $\bar{d}_i$ is the measured round-trip delay of connection i. Fix $\kappa > 0$.

Consider the following system of differential equations:

$$\frac{d}{dt} w_i(t) = -\kappa \frac{d_i}{\bar{d}_i} \frac{s_i}{w_i}, \tag{32}$$

$$s_i = w_i - x_i d_i - p_i \text{ for } i = 1, \cdots, N. \tag{33}$$

Theorem 5 *Let $V(w) = \sum_{i=1}^N \left(\frac{s_i}{w_i}\right)^2$. Then V is a Lyapunov function for the system of differential equations (32)-(33). The unique value minimizing V is a stable point of this system, to which all trajectories converge.*

Proof:

Define

$$J_x = [\frac{\partial x_i}{\partial w_j}, i, j = 1, \cdots, N], \tag{34}$$

$$J_q = [\frac{\partial q_i}{\partial w_j}, i = 1, \cdots, M, j = 1, \cdots, N]. \tag{35}$$

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 6 *The Jacobian $J_x = [\frac{\partial x_i}{\partial w_j}, i, j = 1, \cdots, N]$ of $x(w)$ with respect to w is given by the following expression on the interior point:*

$$J_x = \bar{D}^{-1}(I - X A_B (A_B^T X \bar{D}^{-1} A_B)^{-1} A_B^T \bar{D}^{-1}) \tag{36}$$

where

$$\bar{D} = Diag(d_i + A_{i\cdot}q, i = 1, \cdots, N) \tag{37}$$

$$X = Diag(x_i, i = 1, \cdots, N). \tag{38}$$

Proof: Our starting point is equation (3) which reads

$$x_i[(A_B q_B)_i + d_i] = w_i, i \in B \tag{39}$$

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 $rank(A_B) = |B|$, since otherwise, we can reduce B, to have a full rank and from the proof of theorem 2, the reduced system have same solution as the original system.

With this notation, we have

$$x_i[A_{i\cdot}q + d_i] = w_i \tag{40}$$

$$A^T x = c \tag{41}$$

Taking the partial derivative of (40) 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:

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

In matrix notation, these identities read

$$\bar{D}J_x + XAJ_q = I. \tag{42}$$

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

$$A^T J_x + (A^T\bar{D}^{-1}XA)J_q = A^T\bar{D}^{-1}. \tag{43}$$

Now, (41) implies that $A^T J_x = 0$. Hence,

$$J_q = (A^T \bar{D}^{-1} X A)^{-1} A^T \bar{D}^{-1} \tag{44}$$

Plugging (44) into (42) gives the equation (36) ∎

We now resume the proof of the theorem.

Proof of Theorem:

Let $r_i = \frac{s_i}{w_i}$. 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 $J_r := (\frac{dr_i}{dw_j}, i = 1, \cdots, N, j = 1, \cdots, N)$ is the Jacobian of r. From $J_r = (XDW^{-2} + PW^{-2} - DW^{-1} J_x)$, $\dot{w} = D\bar{D}^{-1} r$, and (36),

$$\frac{d}{dt} V(w(t)) = -\kappa r^T [(XDW^{-2} + PW^{-2} - DW^{-1} J_x) D\bar{D}^{-1}] r. \tag{45}$$

Note that the matrix (45) in the bracket is positive definite. Hence $V(w(t))$ is strictly decreasing in t on the interior points unless $s_j(t) = 0$ for all j. We will argure here that $V(w(t))$ is decreasing even on a boundary point. Assume that at time t, $w(t)$ reaches a boundary point, i.e., at time $t - \epsilon$ w has a bottleneck B and at $t + \epsilon$, $w(t + \epsilon)$ has another bottleneck set $\bar{B}$. Then,

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

Since $V(w(t))$ is right differentiable for all direction and continuous, the expression goes to $\frac{1}{2}(\dot{V_B}(t) + \dot{V_{\bar{B}}})$ where $\dot{V_B}$ right differential of V with bottleneck set B. Hence (46) is sum of two negative number which is negative.

Therefore we have shown that V is a decreasing function of t unless $s_j(t) = 0$ which is unique w minimizing V and the theorem follows from ([22]). ∎

In [20], Kelly et al proposed rate control algorithm converges to *proportionally fair* point. They changes rate as follows:

$$\frac{d}{dt} x_i(t) = \kappa (p_i - x_i(t) A_{i\cdot} \mu_j(t))$$

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 $\mu_j(t)$ play the role of a Lagrange multipliers of the problem P as $\epsilon \to 0$.

Our algorithm, however, controls the window size instead of rate explicitly . The rate is a function of all windows.

$$\frac{d}{dt} w_i(t) = \kappa\left(\frac{p_i}{w_i} + \frac{d_i}{\bar{d}_i} - 1\right)\frac{d_i}{\bar{d}_i}$$

where

$$\bar{d}_i = d_i + \sum_{j \in A_{\cdot i}} q_j.$$

Here, measured delay $\bar{d}_i$ plays the role of implicit feedback. It is the summation of q_j plus d_i. 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, $q = 0$, $\dot{w} = \kappa \frac{p_i}{w_i}$. Increasing rate is a decreasing function of w.

4.2 (p,α)-Proportionally Fair Algorithm

In this subsection, we consider an algorithm that converges to an (p, α)- *proportionally fair* rate vector for $\alpha > 1$. We know that if $s_i^\alpha = w_i - x_i d_i - \frac{p_i}{{x_i}^{\alpha-1}} = 0$ for all i, then the rate vector is (p, α)- *proportionally fair* . We call $\frac{p_i}{{x_i}^{\alpha-1}}$ the "target queue length," since $w_i - x_i d_i$ is the estimated queue length in the network. Note that 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 queue and vice versa.

One unfavorable property of the target queue length function $\frac{p_i}{{x_i}^{\alpha-1}}$ 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 $\frac{p_i}{(x_i+1)^{\alpha-1}}$ instead of $\frac{p_i}{{x_i}^{\alpha-1}}$, 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 $\bar{s}^\alpha = w_i - x_i d_i - \frac{p_i}{x_i+1^{\alpha-1}} = 0$ is

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

Note that $h'_\alpha = \frac{1}{x(x+1)^{\alpha-1}}$ and $\lim_{x\to\infty} h_\alpha = 0$. These observations show that $h_\alpha p$ is increasing concave and nonnegative, and by the claim 5, the solution of (P) with objective function h_α converges to max-min rate vector.

Consider the system of differential equations

$$\frac{d}{dt} w_i = -\kappa x_i s_i u_i \tag{46}$$

where

$$s_i = w_i - x_i d_i - \frac{p_i}{(x_i+1)^{\alpha-1}}, \tag{47}$$

$$u_i = d_i - (\alpha - 1)\frac{p_i}{(x_i+1)^\alpha}. \tag{48}$$

Theorem 6 *If $p_i < \frac{d_i}{\alpha-1}$ for all i, the function $V(w) = \frac{1}{2}\sum_i s_i^2$ is a Lyapunov function for the system of equations (46)-(48). 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 $J_s = (\frac{ds_i}{dw_j} i = 1, \cdots, N, j = 1, \cdots, N)$ is the Jacobian of s with respect to w. The equation (46) can be rewritten in a matrix form as $\dot{w} = -\kappa X U s$ where $U = Diag(u_i, i = 1, \cdots, N)$. 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.

$$\frac{\partial s_i}{\partial w_j} = \delta ij - d_i \frac{\partial x_i}{\partial w_j} + (\alpha - 1)p_i(x_i + 1)^{-\alpha} \frac{\partial x_i}{\partial w_j} \tag{49}$$

or

$$\begin{aligned} J_s &= I - DJ_x + (\alpha - 1)P(X + I)^{-\alpha} J_x && (50) \\ &= I - UJ_x && (51) \\ &= (I - U\bar{D}^{-1}) + U\bar{D}^{-1}XAJ_q && (52) \end{aligned}$$

where $U = Diag(u_i, i = 1, \cdots, N)$ and $J_x = (\frac{\partial x_i}{\partial w_j}, i = 1, \cdots, N, j = 1, \cdots, N)$ Hence

$$\begin{aligned} J_s XU &= XU - (D - (\alpha - 1)PX)J_x XU \\ &= (I - D\bar{D}^{-1})XU + (\alpha - 1)P(X + I)^{-\alpha}\bar{D}^{-1}XU + U\bar{D}^{-1}XAJ_q XU. \end{aligned}$$

Since $\bar{D}^{-1}XAJ_qX$ is positive semidefinite, $(I - D\bar{D}^{-1})$ is diagonal matrix with nonnegative entries and $P(X + I)^{-\alpha}$ is positive definite, $J_s XU$ is positive definite.

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

■

5 CONCLUSIONS

In this paper we have addressed the fundamental question on 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 multiclass closed fluid model. We showed that the flow rates are a well defined function of the window sizes 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 point. We have developed an algorithm which converges to the fair point and proved its convergence using a Lyapunov function.

Our algorithm uses the propagation delay d_i, measured delay $\bar{d}_i$, 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 Richard et al [21] for this problem. The challenging question that remains is the implementability of this protocol.

References

[1] D. Bertsekas and R. Gallager. *Data Networks.* Prentice-Hall, Englwood Cliffs,N.J., 1992.

[2] L.S. Brakmo and L.L. Peterson. Tcp vegas: end to end congestion avoidance on a global internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–80, October 1995.

[3] A. Charny. An algorithm for rate allocation in a packet-switching network with feedback. Master's thesis, MIT, Cambridge,MA, 1994.

[4] D. Chiu and R. Jain. Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. *Computer Networks and ISDN Systems*, 17:1–14, 1989.

[5] D.Mitra. Asymptotically optimal design of congestion control for high speed data network. *IEEE Transaction on Communication*, 40:301–311, February 1992.

[6] D.Mitra and I.Mitrani. Efficient window flow control for high speed data networks with small buffers. *Annals of Operations Research*, 49(1):1–24, March 1994.

[7] C. Douligeris and R. Mazumdar. A Game Theoretic Perspective to Flow Control in Telecommunication Netowk. *Journal of the Franklin Institute*, 1992.

[8] Hahne. E and Gallager. R. Round-robin scheduling for maxmin fairness in data network. *IEEE Journal on Selected Areas in Communications*, 9(7):408–417, September 1991.

[9] K.W. Fendick, D.Mitra, I. Mitrani, M.A. Rodrigues, J.B. Seery, and A Weiss. An approach to high-performance, high-speed data networks. *IEEE Communications Magazine*, pages 74–82, October 1991.

[10] S. Floyd and V. Jacobson. Connection with multiple Congested Gateways in packet-Switched Networks,Part1: One-way Traffic". *ACM Computer Communication Review*, 21(5):30–47, August 1991.

[11] S. Floyd and V. Jacobson. On Traffic Phase Effects in packet Switched Gateways. *Internetworking:Research and Experience*, 3(3):115–156, September 1993.

[12] S. Floyd and V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1(4):397–413, August 1993.

[13] T. Henderson, E. Sahouria, S McCanne, and R. Katz. Improving Fairness of TCP Congestion Avoidance . In *INFOCOM '98*, August 1998.

[14] V. Istratescu. *Fixed Point Theory.* Reidel Rublishing Company, Dordrecht,Holland, 1981.

[15] V. Jacobson. Congestion avoidance and control . *Computer Communication Review*, 18(4):314–29, August 1988.

[16] J.M. Jaffe. Bottleneck flow control. *IEEE Transaction on Communications*, COM-29(7), July 1981.

[17] J.M. Jaffe. Flow Control Power is Nondecentralizable. *IEEE Transactions on Communications*, COM-29(9):1301–1306, September 1981.

[18] R. Jain. Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey . *Computer Networks and ISDN Systems*, 28(13):1723–38, October 1996.

[19] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, pages 33–37, Jan-Feb 1997.

[20] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow price proportional fairness and stability. *Journal of the Operational Research Society*, 1998.

[21] Richard J. La, Jean Walrand, and Venkat Anantharam. Issues in TCP Vegas . available at http://www.path.berkeley.edu/~ hyongla, July 1998.

[22] D. Luenberger. *Introduction to Dynamic System.* John Wiley and Sons, 1979.

[23] A. Mankin. Random Drop Congestion Control. In *Proc. SIGCOMM'90*, pages 1–7, Philadelphia,PA, September 1990.

[24] M.Avriel. *Nonlinear Programming.* Prentice-Hall, Englwood Cliffs,N.J., 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*, pages 926–936, San Francisco, CA, USA, March 1996.

[26] R. Mazumdar, L.G. Mason, and C. Douligeris. Fairness in network optimal flow control:optimality of product forms. *IEEE Transactions on Communications*, 39(5):775–782, May 1991.

[27] R.Jain. Myths about Congestion Management in High Speed Networks. *Internetworking:Research and Experience*, 3:101–113, 1992.

[28] J.B. Rosen. Existence and uniqueness of equilibrium points for concave N-person games. *Econometrica*, 33(3):520–534, 1965.

[29] S. Shenker. A theoretical analysis of feedback flow control. In *ACM SIGCOM '90 Symposium*, Philadelphia, PA, 24-27 Sept 1990.

[30] Z. Zhang and C. Douligeris. Convergence of Synchronous and Asynchronous Greedy Algorithms in a Multiclass Telecommunications Environment. *IEEE Transactions on Communications*, 40(8):1277–1281, August 1992.