Fairness in Network Optimal Flow Control: Optimality of Product Forms

Ravi Mazumdar, *Member, IEEE,* Lorne G. Mason, and Christos Douligeris

Abstract—**In this paper we consider the problem of optimal flow control in a multiclass telecommunications environment where each user (or class) desires to optimize its performance while being fair to the other users (classes). The Nash arbitration scheme from game theory is shown to be a suitable candidate for a fair, optimal operation point in the sense that it satisfies certain axioms of fairness and is pareto optimal. This strategy can be realized by defining the product of individual user performance objectives as the network optimization criterion. This provides the rationale for considering the product of user powers as has been suggested in the literature. For delay constrained traffic, the constrained optimization problem of maximizing the product of user throughputs subject to the constraints leads to a Nash arbitration point. It is shown that these points are unique in throughput space and we also obtain some convexity properties for power and delays with respect to throughputs in a Jackson network.**

I. Introduction

FLOW control has traditionally been used in the context of congestion avoidance in networks. When network resources are limited and meeting the grade of service requirements for each class of traffic is important, then performance-oriented flow control procedures are necessary. Due to the many types of traffic with different and conflicting requirements the problem is one of multiple criteria optimization. This leads to a natural game theoretic framework for the analysis. The use of game theoretic concepts in network optimization has been considered in [5], [12], [16], [18], and [19] where the emphasis was on the characterization of operating points based on game theoretic equilibria. However, the works do not address the issue of the choice of the performance criterion to optimize network performance which is important in the proper apportioning of network resources. In this paper we show how this can be done given the individual objectives.

In earlier work [6], [7] we argued that the network should be operated at pareto-optimal points since mathematically they correspond to equilibria from which any deviation will lead to the degradation in performance of at least one user of class. It can be shown that the pareto-optimal points are characterized by a surface in $(N-1)$ dimensions if there are N individual objectives and thus there are infinitely many solutions in a multiclass environment. We use a criterion of fairness drawn from game theory known as the Nash arbitration scheme to select a unique operating point. This allows us to deal with the isues of optimality and fairness at the same time. Furthermore, it provides us with an abstract framework in which to analyze the difficult concept of fairness with precise mathematical structure.

The organization of the paper is as follows. In Section II we discuss the issue of fairness and performance objectives and present the main result. In Section III we show that user performance based on the power function and throughput maximization subject to constraints on delays satisfy the requirements of Section II. In doing so we show some new convexity results of the inverse of user power and delays in a Jackson-type network. In Section III we offer some concluding remarks.

Paper approved by the Editor for Routing and Switching of the IEEE Communications Society. Manuscript received June 2, 1989; revised May 2, 1990. This work was supported in part by a grant from INRS, by the NSERC under Grant A-1344, and by the National Science Foundation under Grant CDR-84-21402. This paper was presented at the SBT/IEEE International Telecommunications Symposium, Rio de Janeiro, Brazil, September 3–6, 1990.

R. Mazumdar and L. G. Mason are with INRS-Télécommunications, Université du Québec, Ile-des-Soeurs, P. Q., Canada H3E 1H6.

C. Douligeris is with the Department of Electrical Engineering, University of Miami, Coral Gables, FL 33124.

IEEE Log Number 9144540.

II. Performance Measures: Fairness and Optimality

The aim of any network performance optimization procedure must be the better utilization of network resources while providing a satisfactory level of performance for each user or class. In the emerging integrated network environment the users or classes can be differential based on the grade of service required. Thus, it makes little sense to maximize an overall network performance measure without regard to the actual performance of each user of class. For packet switching traffic the important performance measures are throughput and delay while blocking is the important measure for connection-oriented traffic. In this paper we consider performance objectives most relevant to the packet switched environment although the main results do not depend on the particular situation.

Game theory provides a natural framework for the analysis of the problem. This is not just an artifact, the advantage is that now we have a precise mathematical framework. This allows us to address the important issues of fairness and proper operating points for the network. References [1] and [17] are good references on game theory.

In a game theoretic setting there are two inherently different types of situations: cooperative and noncooperative games. The noncooperative game framework is one in which every class or user acts individually to optimize its performance measure without regard to the performance of other classes. Such a procedure leads to a Nash equilibrium point in the network [12]. This situation is important when the users act based only on local information [2], [8]. However, if the users

0090-6778/91/0500-0775$01.00

play a cooperative game then the performance of each class or user may be made better than the performance achieved at the Nash equilibrium. This is because the Nash equilibrium is pareto inefficient under certain conditions [9]. Hence, it is desirable to operate the network at pareto-optimal points. With the cooperative framework as the basis, we can then study the important issue of fairness.

The issue of fairness has been an important component in the design of optimal flow control schemes since it has been shown that there exist situations where a given scheme might optimize network throughput while denying access to a particular (or a set of) user(s) [10]. However, fairness is difficult to quantify in the absence of a proper framework. Loosely speaking, fairness can be thought of as a situation in which no individual class or user is denied access to the network or overly penalized. The Nash equilibrium or competitive equilibrium can be shown (for the case of the power criterion) to be a point where no user is denied access to the network and in particular if the performance objectives are the same then it corresponds to equal throughputs for each class. From the above discussion it can be considered a fair operating point and in fact is the essence of the fair scheme proposed in [2]. However, since pareto-optimal points can provide higher user performance in general than Nash points, the key is to choose the appropriate pareto-optimal point (in general there are infinitely many) such that it guarantees that user objectives are met.

In [10], Gerla and Staskausas define a notion of *optimal fairness* in which total throughput is maximized subject to the network capacity being fairly utilized. A scheme which provides for equal sharing when the demands exceed capacity is then suggested as a fair scheme. From a game-theoretic standpoint, such a point is not special. Moreover, the tradeoff between throughput and delay is not taken into account. Several other ad hoc schemes might be proposed based on the ratios of individual demands or the precise nature of the individual performance objectives.

We now introduce a notion of fairness drawn from the cooperative game framework which has a precise mathematical interpretation which subsumes the usual assumptions as to what constitutes fairness. The most important outcome is that it leads to the optimization of a unique performance measure which is characterized completely by the individual performance measures.

The key notion of a fair strategy in cooperative game theory is the notion of the *Nash arbitration strategy* [20]. In order for a strategy to be a Nash arbitration strategy it should satisfy the axioms of fairness given below. See [17] for a discussion of the Nash arbitration scheme. It is important to note the difference between the Nash arbitration strategy or scheme and the Nash equilibrium. The former corresponds to the cooperative situation while the latter is the competitive equilibrium. The Nash or competitive equilibrium point is a point at which no user may deviate from in order to improve its performance given that all the players adopt the same strategy. This corresponds to a player optimizing its performance without regard to the performance of the others before. A pareto equilibrium corresponds to the situation where no individual can improve its performance without affecting at least one user adversely.

In order to state the axioms we first introduce the mathematical framework.

Consider a cooperative game of N players (users). Let each individual player i have an objective function $f_i(x) : X \rightarrow \mathcal{R}$ where X is a convex, closed, and bounded set of $\mathcal{R}^N$. From the point of view of communications networks X will denote the space of throughputs. Let $u^* = [u_1^*, u_2^*, \cdots, u_N^*]$ where $u_i^* = f_i(x^*)$ for some $x^* \in X$ denote a common agreement point which all the players agree to as a starting point for the game. In general u^* can be thought of as the vector of individual user performances which the user would like to at least achieve if they enter the game. Let $[U, u^*]$ denote the game defined on X with initial agreement point u^* where U denotes the image of the set X under $f(\cdot)$, i.e., $f(X) = U$. Let $F[\cdot, u^*] : U \rightarrow U$ be an arbitration strategy. Then F is said to be a Nash arbitration strategy if it satisfies the four axioms below.

1) Let $\phi(u) = u'$ where $u_i' = a_i u_i + b_i$ for $i = 1, 2, \cdots, N$ and $a_i > 0$, b_i are arbitrary constants. Then
$$F[\phi(U), \phi(u^*)] = \phi(F[U, u^*]).$$
This states that the operating point in the space of strategies is invariant with respect to linear utility transformations.

2) The arbitration scheme must satisfy
$$(F[U, u^*])_i \geq u_i^* \qquad \text{for } i = 1, 2, \cdots, N$$
and furthermore there exists no $u \in X$ such that $u_i \geq (F[U, u^*])_i$ for all $i = 1, 2, 3, N$. This is the notion of pareto optimality of the arbitrated solution.

3) Let $[U_1, u^*]$ *and* $[U_2, u^*]$, be two games with the same initial agreement point such that:

i) $U_1 \subset U_2$

ii) $F[U_2, u^*] \in U_1$.

Then $F[U_1, u^*] = F[U_2, u^*]$.

This is called the independence of irrelevant alternatives axiom. This states that the Nash arbitration scheme of a game with a larger set of strategies is the same as that of the smaller game if the arbitration point is a valid point for the smaller game. The additional strategies are superfluous.

4) Let U be symmetrical with respect to a subset $J \subseteq \{1, 2, 3, \cdots, N\}$ of indexes (i.e., let $i, j \in J$ and $i < j$, then $\{u_1, u_2, \cdots, u_{i-1}, u_j, u_{i+1}, \cdots, u_{j-1}, u_i, u_{j+1}, \cdots, u_N\} \in U$.

If $u_i^* = u_j^*$, then $(F[U, u^*])_i = (F[U, u^*])_j$ for $i, j \in J$.

This is the axiom of symmetry which says that if the set of utilities is symmetric then for any two players if the initial agreement point corresponds to equal performance then their arbitrated values are equal.

Remark: Note that the above axioms guarantee that no user (or class) is denied access to the network if $u^* = 0$ (provided superior points exist) and the arbitrated solution is at least as good as the Nash equilibrium if u^* is taken to be the Nash equilibrium. Thus, in particular, the axioms imply that a Nash arbitration strategy for the network in which the users have the same performance objectives will correspond to equal sharing if the set of admissible throughputs is symmetric and if the initial agreement point is chosen to be one which corresponds to equal throughputs by Axiom 4.

The following theorem (due to Stefanescu and Stefanescu [21]) characterizes the Nash arbitration scheme.

Theorem 1 (Nash Arbitration Scheme): Let $f_i : X \rightarrow \mathcal{R}$ $i = 1, 2, \cdots, N$ be concave, upper bounded, functions defined on X a convex, closed, and bounded set of $\mathcal{R}^N$.

Let $U = \{u \in \mathcal{R}^N : \exists\, x \in X \text{ s.t. } u \leq f(x)\}$ and $X(u) = \{x : u \leq f(x)\}$ and $X_0 = X(u^*)$.

Then the Nash arbitration scheme is given by the point which maximizes the unique function

$$V(x) = \prod_{1}^{N} (f_i(x) - u_i^*) \qquad \text{over } X_0$$

if X_0 contains vectors x which results in the user objectives strictly superior to u^*. If the vectors in X_0 have the property that there exist $x \in X_0$ such that only k of the individual objectives are superior to the corresponding elements of u^*, then the unique function is taken as the product of the individual objectives for which there exist superior solutions. The remaining $(n - k)$ components of u^* are the user objectives at the Nash arbitration point.

Remark: It is important to note that the solution in general depends on the initial agreement point. The point where $V(x)$ is maximized is defined as the fair network optimal operating point.

The use of the power function as the ratio of the average throughput over the average delay has been used in the context of flow control for some time [14], [11], [13]. In fact, it has been noted that the product of powers is a more appropriate optimization criterion [2] since the overall network power was not found to be suitable. In the following section we shall show that the inverse of the power function satisfies the assumptions of the theorem and thus the above result justifies the use of the product of powers as a network optimization criterion.

Before concluding this section it is important to note that the Nash arbitration strategy is not the only *fair* arbitration scheme possible. In fact, standard criticisms of the Nash scheme (see Luce and Raiffa [17] for a complete discussion) led to the development of other arbitration schemes due to Raiffa [17] and Thompson [3]. However, it can be shown that these other schemes correspond to Nash arbitration schemes for performance objectives obtained by linear transformations of the original objectives [3] and thus we restrict ourselves to the Nash arbitration scheme.

In the next section, we define some additional performance objectives for the design of optimal, fair flow control schemes. These will be shown to satisfy the hypothesis of the theorem and thus the existence of the Nash arbitration scheme. Moreover, we shall show that the optimization results in unique points in the throughput space.

III. Optimal Fair Solutions: Existence and Uniqueness

In this section we describe and analyze three performance measures for the design of optimal, fair flow control schemes in a packet switched integrated telecommunications environment. The first criterion is the product of user powers (PPC) where power is defined as the ratio of the average throughput over the average delay of a particular user or class. It is shown that the stationary point for the PPC is a Nash arbitration scheme and gives rise to a unique set of user throughputs. The second formulation is what we term the modified throughput/delay criterion (MTD) which is a generalization of a criterion originally proposed by Lazar [16]. The final criterion is based on a barrier function approach in order to circumvent the constrained nonlinear optimization which results in the MTD formulation.

A. Product of Power Criterion

The product of powers (PPC) as a network performance criterion has been proposed in the context of performance-oriented flow control for single class packet switched networks by Bharathkumar and Jaffe [2]. This was due to the fact that the overall network power was found to be unsuitable as it was deemed as lacking fairness properties. It was also noted that there could be difficulties associated with the nonconcavity of the user power function. The results reported here show that the maximization of the PPC results in a Nash arbitration scheme and moreover the Nash arbitration scheme is unique in the space of throughputs, a strong result in light of the nonconcavity.

When working with the user power function Theorem 1 cannot be directly applied to show the existence of the Nash arbitration scheme due to the nonconcavity of the individual user power. We show, however, that the inverse of user power is convex with respect to the throughputs and using this property we show the existence of a Nash arbitration scheme. We then show that this result is also unique in the space of throughputs.

Let $S = [S_1, S_2, \cdots, S_N]^T$ denote the average throughputs for the N players in the network. We assume that the network is modeled as a Jackson network of $M/M/1$ queues with loop-free routing. We also assume that there are L links with link capacities $C = [c_1, c_2, \cdots, c_L]^T$.

Let $\Delta = [\Delta_1, \Delta_2, \cdots, \Delta_N]^T$ denote the vector of corresponding user delays.

Let $P_i = \frac{S_i}{\Delta_i}$ $i = 1, 2, \cdots, N$ denote the power function of user i which is defined on the set of admissible throughputs

$$U = \{S \geq 0 : 0 \leq \gamma_l < c_l; \qquad l = 1, 2, L\}$$

where γ_l denotes the total throughput on link l.

Lemma 2.1: For a Jackson network with loop-free routing the inverse of the power of user i, $i = 1, 2, \cdots, N$ defined by

$$P_i^{-1} = \frac{\Delta_i}{S_i} \tag{1}$$

is convex in the space of throughputs, i.e., P_i^{-1} is a convex function of S.

Proof: To prove the assertion we first decompose the expression for inverse power in terms of the link contributions. This is because the average user delay is additive over links. Thus

$$P_i^{-1} = \sum_l P_{il}^{-1} = \sum_l \frac{\Delta_{il}}{S_i}$$

where Δ_{il} is the contribution to the delay of user i by link l.

We now establish the convexity of P_{il}^{-1} by showing that the Hessian matrix

$$H_{il} = \left[\frac{\partial^2 P_{il}^{-1}}{\partial S_j \partial S_k}\right]$$

is positive semidefinite.

We first consider the case of fixed routing in the network. Then the following expressions are easily shown:

$$\frac{\partial^2 \Delta_{il}}{\partial S_k \partial S_j} = \frac{2\alpha}{(c_l - \gamma_l)^3}\delta_{lj}\delta_{lk} \tag{2}$$

$$\frac{\partial \Delta_{il}}{\partial S_k} = \frac{\alpha}{(c_l - \gamma_l)^2}\delta_{lk} \tag{3}$$

$$\Delta_{il} = \frac{\alpha}{(c_l - \gamma_l)} \tag{4}$$

where α is the mean packet length and δ_{ij} denotes the Kronecker delta function. Hence, by direct calculation we obtain

$$\frac{\partial^2 P_{il}^{-1}}{\partial S_k \partial S_j} = \frac{1}{S_i}\frac{\partial^2 \Delta_{il}}{\partial S_k \partial S_j} - \frac{1}{S_i^2}\frac{\partial \Delta_{il}}{\partial S_j}\delta_{ik} - \frac{1}{S_i^2}\frac{\partial \Delta_{il}}{\partial S_k}\delta_{ij} + \frac{2}{S_i^3}\Delta_{il}\delta_{ij}\delta_{ik}. \tag{5}$$

For the purpose of notational simplicity we suppress the user index. This is legitimate since interchanging rows and corresponding columns of a matrix does not alter its character (i.e., positive semidefiniteness, etc.). Thus, the $(1,1)$ element of H_{il} can be considered as $\frac{\partial^2 P_{il}^{-1}}{\partial^2 S_i}$ for any $i = 1, 2, \cdots, N$.

Let M_1, M_2, M_k denote the leading principal minors of dimensions $1 \times 1, 2 \times 2, \cdots, k \times k$. In general, $k \le N$ since all the different classes need not share the given link. Then it is straightforward to show that

$$\det M_1 = \frac{2\alpha}{S^3(c_l - \gamma_l)^3}\left[S^2 - S(c_l - \gamma_l) + (c_l - \gamma_l)^2\right] \ge \frac{2\alpha}{S^3(c_l - \gamma_l)^3}[S - (c_l - \gamma_l)]^2 \ge 0$$

for all feasible throughputs, i.e., $S \ge 0$, $c_l - \gamma_l > 0$.

The second principal minor can be shown to be

$$M_2 = \frac{2\alpha}{S^3(c_l - \gamma_l)^3} \cdot \begin{bmatrix} S^2 - S(c_l - \gamma_l) + (c_l - \gamma_l)^2 & S^2 - \frac{S}{2}(c_l - \gamma_l) \\ S^2 - \frac{S}{2}(c_l - \gamma_l) & S^2 \end{bmatrix}$$

and hence

$$\det M_2 = \frac{3}{2S(c_l - \gamma_l)} > 0.$$

It can be shown that the higher order leading principal minors M_k for $k = 3, 4, \cdots, N$ have the property that $\det M_k = 0$. This follows from the fact that the Hessian matrix has the form

$$\begin{bmatrix} a-2b+c & a-b & a-b & \cdots & a-b & 0 & \cdots & 0 \\ a-b & a & a & \cdots & a & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ a-b & a & a & a & a & 0 & & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

and hence for $k \ge 3$ the minors contain repeated rows and thus are singular. The 0's in the matrix arise if class j does not use link l which user i takes. From above it also follows that all the remaining principal minors have determinant 0.

Hence, it follows that H_{il} is positive semidefinite which demonstrates the convexity of P_{il}^{-1}. Since P_i^{-1} is the sum over all links l used by i, it too is convex.

The convexity of P_i^{-1} in the case with random loop-free routing follows immediately since the corresponding Hessian is convex combination of the Hessian with fixed routing.

Remark: The loop-free assumption is reasonable in communication networks where the routing is usually feedforward.

We now use the above result to show the existence of a Nash arbitration scheme for the case of PPC.

Theorem 2.2: Consider a Jackson network with loop-free routing with N users. Let the performance objective of each user be the power function defined by

$$P_i(S) = \frac{S_i}{\Delta_i(S)}; \qquad i = 1, 2, \cdots, N$$

where S_i is the average throughput of user i, $\Delta_i(S)$ represents the corresponding average delay, and S the vector of user throughputs.

The flow control scheme which maximizes the product of the user powers (PPC) is an optimal, fair flow control scheme in the sense that it corresponds to a Nash arbitration scheme for $-P_i^{-1}$ and given by

$$S^* = \arg\max \prod_{i=1}^{N} P_i(S). \tag{6}$$

Moreover, S^* is unique.

Proof: Note, from the previous lemma, $P_i^{-1}(S)$ is convex for each i and hence $-P_i^{-1}(S)$ is concave. Working in the inverse power space precludes us from choosing the point $[0, 0, \cdots, 0]^T$ as the initial agreement point. Hence, we need to choose an initial agreement point u^* in the inverse power space which is achievable in the set of feasible throughputs X. Before proceeding to show that S^* indeed corresponds to a Nash arbitration scheme we first show that S^* is pareto optimal and unique.

Consider the functions $P_i(S)$ and $\prod_1^N P_i(S)$. Then since $P_i(S)$ is defined on X which is compact and convex, $\prod_1^N P_i(S)$ is continuous, and hence achieves its maximum on X. Since $P_i(S)$ is zero on the boundary of X it implies that the maximum is achieved in the interior of X. Hence, the necessary condition that S^* satisfies is

$$\nabla_S \prod_1^N P_i(S)|_{S^*} = 0. \tag{7}$$

Rewriting this in matrix form gives

$$\begin{bmatrix} \frac{\partial P_1}{\partial S_1} & \frac{\partial P_2}{\partial S_1} & \cdots\cdots & \frac{\partial P_N}{\partial S_1} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \frac{\partial P_1}{\partial S_i} & \frac{\partial P_2}{\partial S_i} & \cdots\cdots & \frac{\partial P_N}{\partial S_i} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \frac{\partial P_1}{\partial S_N} & \frac{\partial P_2}{\partial S_N} & \cdots\cdots & \frac{\partial P_N}{\partial S_N} \end{bmatrix} \begin{bmatrix} \prod_{j\neq 1} P_j(S) \\ \cdots \\ \prod_{j\neq i} P_j(S) \\ \cdots \\ \prod_{j\neq N} P_j(S) \end{bmatrix} = 0$$

$$\text{at } S = S^*. \tag{8}$$

Let $J(S)$ denote the matrix above. Then from the observation that $P_i(S^*) \neq 0$ for all i, this implies that the matrix $J(S^*)$ is singular. But the matrix $J(S)$ is the transpose of the Jacobian matrix for $[P_1(S), P_2(S), \cdots, P_N(S)]^T$ and, hence, it implies that $\text{Det}|J(S^*)|$ is zero. But this is the necessary condition for a point to be pareto optimal [22]. But from the fact that any (see the following) deviations from S^* results in at least one player with lower power it is also sufficient and thus S^* is the pareto optimal point.

We now compare this point to the Nash equilibrium point. That the Nash equilibrium exists follows from the fact that the functions $P_i(S)$ are concave in their own throughputs, i.e., w.r.t. S_i (see Rosen [23]). The Nash equilibrium point for this case corresponds to the point S at which

$$\left.\frac{\partial P_i(S)}{\partial S_i}\right|_{\overline{S}} = 0. \tag{9}$$

Hence, it is readily seen that the Nash equilibrium is not pareto optimal and hence is pareto inefficient (see [9]).

We now show the uniqueness of the point S^* in the space of admissible throughputs.

First note that from above it follows that the stationary point results in nonzero throughputs for each user. Define $\prod(S) = \prod_{i=1}^N P_i(S)$. Then let S^* denote that stationary point of $\prod(S)$ then it is easy to show that the necessary condition that S^* maximize $\prod(S)$ is given by

$$\frac{1}{S_i^*} = \frac{1}{\prod(S^*)} \frac{\partial \prod}{\partial S_i}(S^*) \qquad i = 1, 2, \cdots, N. \tag{10}$$

Consider a perturbation of the point S^* given by

$$S = S^* + K\epsilon$$

for some feasible direction in the space of throughputs, i.e.,

$$S_i = S_i^* + k_i\epsilon, \qquad i = 1, 2, \cdots, N.$$

Then the product of throughputs is given by $\prod_1^N (S_i)$. Normalizing this at the point S^* gives the normalized product of throughputs as

$$\prod_1^N \left(1 + \frac{k_i}{S_i^*}\epsilon\right) = \prod_1^N \left(1 + k_i\epsilon \sum_{s\cap i} \left[\frac{\sum_{l\cap s\cap i} \frac{\alpha}{e_l^2}}{\sum_{l\cap s} \frac{\alpha}{e_l}}\right]\right) \tag{11}$$

where $e_l = c_l - \gamma_l^*$ the residual flow in the link under S^* and $l \cap s \cap i$ denotes the set {users s and i which use link l}.

Similarly the product of the mean delays is given by

$$\prod_1^N \left(\sum_{l\cap i} \frac{\alpha}{e_l - \sum_{l\cap s\cap i} k_s\epsilon}\right).$$

Upon normalization with respect to the delays at the stationary point and rearrangement this can be written as

$$\prod_1^N \left(1 + \frac{\sum_{l\cap i} \sum_{l\cap s\cap i} \frac{\alpha k_s\epsilon}{e_l(e_l - \sum_{s\cap l\cap i} k_s\epsilon)}}{\sum_{l\cap i} \frac{\alpha}{e_l}}\right). \tag{12}$$

Now it is easy to show that this is greater than (strict if at least one of the k_s is nonzero)

$$\prod_1^N \left(1 + \frac{\sum_{l\cap i} \sum_{l\cap s\cap i} \frac{\alpha k_s\epsilon}{e_l^2}}{\sum_{l\cap i} \frac{\alpha}{e_l}}\right). \tag{13}$$

By comparing (11) to (13) it can be easily seen that the normalized product of perturbed throughputs is less than the normalized product of perturbed delays implying that the perturbed PPC normalized around the stationary point is < 1. This implies that any perturbation of the stationary point is not optimal establishing the uniqueness of the maximizing point in the throughput space.

Having established the uniqueness of the point S^* we now show that it corresponds to a Nash arbitration point for the negative inverse power.

Take as the initial agreement point u^* the point where $u_i^* = -aP_i^{-1}(S^*)$ $i = 1, 2, \cdots, N$ for $a > 1$ sufficiently large. Then u^* is a valid initial agreement point for a game played by N players with $-P_i^{-1}(S)$ as individual objectives. This is because if a is sufficiently large, then $-aP_i^{-1}(S^*) < \max_S -P_i^{-1}(S)$ and the maximum of the individual negative inverse powers exists since the functions are concave over a convex, compact domain of feasible throughputs X and attain the value $-\infty$ on the boundary of X. Now in order to apply Theorem 1 to the concave, upperbounded (because of the comment above) functions $-P_i^{-1}(S)$ we need to show the convexity and compactness of the set

$$X_0(u^*) = \{S \,:\, u \in U \text{ s.t. } u_i \geq u_i^* \qquad i = 1, 2, \cdots, N\}$$

where

$$U = \{u \,:\, \exists\, S \in X \text{ s.t. } u_i \leq -P_i^{-1}(S), \; i = 1, 2, \cdots, N\}.$$

Note that by our choice of u^*, $X_0(u^*)$ is nonempty. Choose any arbitrary points u^1 *and* u^2 in U. Then to show that $X_0(u^*)$ is convex it is enough to show that U is convex. To show that U is convex we need to show that $u^3 = c\,u^1 + (1-c)\,u^2$, for $0 \leq c \leq 1$ is in U.

Let S^1 *and* S^2 be two throughput vectors corresponding to u^1 and u^2, respectively. Then from the definition of U we have

$$c(-P_i^{-1}(S^1)) + (1-c)(-P_i^{-1}(S^2)) \geq u_i^3$$

and from the concavity of $-P_i^{-1}(S)$ over X we have

$$-P_i^{-1}(S^3) = -P_i^{-1}(cS^1 + (1-c)S^2) \geq -cP_i^{-1}(S^1) - (1-c)P_i^{-1}(S^2) \geq u_i^3.$$

The convexity of X implies that S^3 is a valid throughput and hence u^3 belongs to U. Hence, the set $X_0(u^*)$ is convex. Compactness follows from the fact that the set is closed and bounded.

Hence, applying Theorem 1 to the user functions $-P_i^{-1}(S)$, $i = 1, 2, \cdots, N$ with initial agreement point u^*, the Nash arbitration scheme exists and is the point which maximizes $\prod_1^N (-P_i^{-1}(S) - u_i^*)$ over $X_0(u^*)$. The necessary conditions for this are

$$J(S) \begin{bmatrix} \frac{\prod_{j\neq 1}(-P_j^{-1}(S) - u_j^*)}{P_1(S)^2} \\ \cdots \\ \frac{\prod_{j\neq i}(-P_j^{-1}(S) - u_j^*)}{P_i(S)^2} \\ \cdots \\ \frac{\prod_{j\neq N}(-P_j^{-1}(S) - u_j^*)}{P_N(S)^2} \end{bmatrix} = 0 \qquad (14)$$

at the stationary point $J(S)$ is the matrix defined above and corresponds to the transpose of the Jacobian for the vector of powers. Note that $P_i(S) \neq 0$ at the stationary point since $X_0(u^*)$ excludes the point $[0, 0, \cdots, 0]$.

Multiply the vector on the LHS of (14) by $(-1)^{N-1}$ $(\prod_1^N P_i(S))^2$, then the LHS of (14) evaluated at the point S^* can be written as

$$J(S^*) \begin{bmatrix} \left[\prod_{j\neq 1} P_j(S^*)\right](1-a)^{N-1} \\ \cdot\cdot \\ \left[\prod_{j\neq i} P_j(S^*)\right](1-a)^{N-1} \\ \cdot\cdot \\ \left[\prod_{j\neq N} P_j(S^*)\right](1-a)^{N-1} \end{bmatrix}.$$

But from the definition of S^* we see that the vector $\left[\prod_{j\neq 1} P_j(S^*), \cdots, \prod_{j\neq N} P_j(S^*)\right]^T$ belongs to the null space of $J(S^*)$ and hence S^* satisfies the necessary condition for it to be the stationary point of $\prod_1^N (-P_i^{-1}(S) - u_i^*)$. The concavity of the functions $-P_i^{-1}(S) - u_i^*$ implies that the condition is also sufficient and hence S^* is the Nash arbitration scheme for the negative inverse powers and the proof is done.

Remarks: In the proof of the theorem we noted that the Nash equilibrium point is pareto inefficient. This implies that there exist points at which the user powers are superior to those achieved at the Nash equilibrium. Hence, let $u^* = \text{col}[u_1^*, u_2^*, \cdot, \cdot, u_N^*]$ denote the Nash equilibrium point. Then the point

$$S^{**} = \arg\max \prod_{i=1}^{N} (P_i(S) - u_i^*)$$

will correspond to the pareto point strictly superior to the Nash equilibrium point, i.e., it corresponds to the case where the Nash equilibrium point is chosen as the initial agreement point. Note S^{**} is unique.

The lack of concavity of the user power function presents difficulties in concluding that the point which maximizes the PPC is a Nash arbitration scheme for the power criterion. This is due to the difficulty of showing that the set U of allowable powers is convex and compact. However, several nontrivial examples have been worked out which show that the set U is in fact convex and compact. This leads us to conjecture that the maximization point of the PPC is in fact the Nash arbitration scheme for the power function with $[0, 0, \cdots, 0]$ as the initial agreement point. For the case of $M/M/1$ queues the convexity and compactness of the set of achievable powers over the set of feasible throughputs has been shown in [6].

The above results provide the rationalization for the use of the PPC. An advantage in using the power criterion is that the optimization problem is unconstrained since the maximum is always attained in the interior of the set of feasible throughputs. However, one serious drawback is that in real applications there are usually constraints on the allowable user delays which power does not take into account. Power introduces weak constraints in that it is inversely proportional to the delays and thus the maximum of the product may be achieved at throughputs which violate delay constraints. What is needed is a network performance criterion which explicitly takes into account the delay constraints while ensuring that the resulting solution is fair. We now focus attention on defining such a network criterion.

B. Modified Throughput/Delay Criterion

In the context of flow control of queueing networks, Lazar [16] introduced the throughput/delay criterion which seeks to maximize the average throughput subject to bounds on the average delay. In the context of optimum, fair schemes we can generalize this notion to define a modified throughput/delay (MTD) criterion whose optimum is achieved by a Nash arbitration scheme. This will be the cooperative equilibrium when each user's objective is the maximization of its own throughput subject to constraints on the delay.

Let S_i denote the user throughputs with $\Delta_i(S)$ the corresponding delays. Let $\delta = [\delta_1, \delta_2, \cdots, \delta_N]^T$ denote the vector of constraints on the delay, i.e., $\Delta_i(S) < \delta_i$ for $i = 1, 2, \cdots, N$.

Lemma 2.3: The user delays $\Delta_i(S)$ for $i = 1, 2, \cdots, N$ are convex in S and increasing in each S_i.

Proof: This assertion can be proved in an analogous manner to Lemma 2.1. This is done by decomposing the flows on links assuming fixed routing and using the fact that delays are additive on links. It can then be shown that the delay component along each link in convex by showing that the Hessian is nonnegative definite. From this and the fact that the Hessian in the case with random loop-free routing is a convex combination of the Hessian in a fixed case it will follow that the Hessian matrix is nonnegative definite. Thus, convexity is proved. The increasing property follows from the fact that the gradient is positive.

Lemma 2.4: The set of admissible throughputs for the constrained problem

$$U = \{S \text{ feasible} : \Delta_i(S) \leq \delta_i, \quad i = 1, 2, \cdots, N\}$$

is a convex and compact set.

Proof: The convexity of U follows from the convexity of $\Delta_i(S)$. That the set is closed and bounded (compact) follows from the fact that $\Delta_i(S)$ is increasing in S_i and thus there exists an S_i at which the constraint is achieved if the constraint is active if not S_i is bounded by the maximum feasible flow for stability.

Theorem 2.5: Define the modified throughput/delay function

$$\mathrm{MTD}(S) = \prod_{i=1}^{N} S_i \tag{11}$$

subject to $\Delta(S) \leq \delta$ and S being feasible.

Then the MTD is the unique function that determines the Nash arbitration scheme for user throughput maximization subject to constraints on the average user delay. Moreover, it results in a unique set of user throughputs.

Note $\Delta \leq \delta$ is to be understood as componentwise.

Proof: We now take the initial agreement point $u^* = [0,0,0,\cdots,0]^T$ which implies that every user desires nonzero throughput. Then the proof follows directly from Theorem 1 since S_i is linear and hence concave. It is upperbounded since the set U is compact. Also since U is convex from Theorem 1, $\prod_1^N S_i$ is the unique function that determines the Nash arbitration scheme. The uniqueness of the maximizing point in the space of throughputs follows since MTD(S) is a multilinear function defined on a convex, compact set and thus achieves its maximum at a unique point.

C. Discussion

The drawback of the MTD criterion is that the Nash arbitration strategy is determined by solving a nonlinear program with convex constraints. Thus, although the criterion captures the essence of the flow control problem, in practice for large networks the nonlinear program can be quite cumbersome. To circumvent this difficulty we present an approximate formulation which is exact in the limit such that the modified problem results in an unconstrained problem. (Note there are constraints due to feasibility of flows but the maximum of products is achieved in the interior of the feasible set and thus the feasibility constraints will be inactive.) This is based on a barrier function approach to solve the MTD problem. In this connection, it should be pointed out that Giessler *et al.* [11] suggested using the generalized power criterion where the throughputs are weighted by a factor of the form S_i^{β}. However, there is no reasonable way to guarantee that a particular choice of the weighting factor will satisfy the delay constraints. It is known that there exist (throughput, delay) pairs which are feasible but cannot be achieved by choice of weighting factors for each user, i.e., the weak constraint problem still exists in the generalized power case.

Hence, define the barrier function as

$$BF(S) = \prod_{1}^{N} S_i \left(1 - \frac{\Delta_i(S)}{\delta_i}\right)^{\beta_i}, \qquad 0 \leq \beta_i \leq 1 \tag{15}$$

where the $\beta_i' s$ are weighting factors chosen such that the sequence of $\beta_i' s$ are monotone decreasing and converge to zero. Then it can be shown that the solution converges to the solution of the MTD criterion [24]. The advantage is that we now have an unconstrained problem. We note that other formulations based on the penalty function can be also used to approximate the MTD solutions.

IV. Conclusion

In this paper we have presented a precise mathematical formulation and characterization of the issue of the design criteria for network optimal flow control and the related issue of fairness. By using the game theoretic framework, we have identified the Nash arbitration scheme as a desirable optimal, fair operating point for the individual users. Furthermore, the strategy can be obtained by only knowing the individual user performance criteria. We have provided a proof of why the product of powers is indeed a reasonable design criterion and shown some new convexity properties of the power function and user delay functions. We have also given a more appropriate criterion in the context of packet switched networks which solves the optimum, fair flow control issue when there are constraints on user delays.

These result could be thought of as the first concrete attempt at providing a mathematical basis for optimal flow control and fairness in the network context. An important issue which arises is the design of decentralized algorithms to achieve these operating points and the extension of these ideas to the general environment where there is mixed type of traffic.

References

[1] T. Basar and G. T. Oldser, *Dynamic Noncooperative Game Theory.* New York: Academic, 1982.

[2] K. Bharathkumar and J. M. Jaffe, "A new approach to performance oriented flow control," *IEEE Trans. Commun.,* vol. COM-29, pp. 427–435, Apr. 1981.

[3] X. Cao, "Preference functions and bargaining solutions," in *Proc. 21st CDC,* Orlando, FL, pp. 164–171, Dec. 1982.

[4] D. Cansever, "Decentralized algorithms for flow control in networks," in *Proc. 25th CDC,* Athens, Greece, Dec. 1986.

[5] C. Douligeris and R. Mazumdar, "An approach to flow control in an integrated environment," CTR Tech. Rep., Columbia Univ., CUCTR-TR 50, 1987.

[6] ——, "On pareto optimal flow control in an integrated environment," in *Proc. 25th Allerton Conf.,* Univ. Illinois, Urbana, Oct. 1987.

[7] ——, "More on pareto optimal flow control," in *Proc. 26th Allerton Conf.,* Univ. Illinois, Urbana, Oct. 1987.

[8] ——, "User optimal flow control in an integrated environment," in *Proc. Indo–U.S. Workshop on Syst. Signals,* Bangalore, India, Jan. 1988.

[9] P. Dubey, "Inefficiency of Nash equilibria," *Math. Oper. Res.,* vol. 11, no. 1, pp. 1–8, 1986.

[10] M. Gerla and M. Staskausas, "Fairness in flow controlled networks," *J. Telecommun. Networks,* pp. 29–38, 1982.

[11] A. Giessler, J. Hanle, A. Konig, and E. Pade, "Free buffer allocation—An investigation by simulation," *Comput. Networks,* vol. 1, no. 3, pp. 199–204, 1978.

[12] M. T. Hsiao and A. Lazar, "A game theoretic approach to optimal decentralized flow control of markovian queueing networks with multiple controllers," in *Proc. Performance 87,* Brussels, Belgium, Dec. 1987.

[13] J. M. Jaffe, "Flow control power is nondecentralizable," *IEEE Trans. Commun.,* vol. COM-29, pp. 1301–1306, Sept. 1981.

[14] L. Kleinrock, "Power and deterministic rules of thumb for probabilistic problems in computer communications," in *Proc. ICC,* June 1979, pp. 43.1.1–43.1.10.

[15] L. Kleinrock and M. Gerla, "Flow control—A comparative survey," *IEEE Trans. Commun.,* vol. COM-28, pp. 553–574, 1980.

[16] A. Lazar, "Optimal flow control of a class of queueing networks in equilibrium," *IEEE Trans. Automat. Contr.,* vol. AC-28, pp. 1001–1007, 1983.

[17] R. D. Luce and H. Raiffa, *Games and Decisions.* New York: Wiley, 1957.

[18] L. Mason, "Equilibrium flows, routing patterns and algorithms for store-and-forward networks," *J. Large Scale Syst.,* vol. 8, pp. 187–209, 1985.

[19] L. Mason and A. Girard, "Control techniques and performance models in circuit switched networks," in *Proc. IEEE CDC,* Orlando, FL, Dec. 1982.

[20] J. Nash, "The bargaining problem," *Econometrica,* vol. 18, pp. 155–162, 1950.

[21] A. Stefanescu and M. W. Stefanescu, "The arbitrated solution for multiobjective convex programming," *Rev. Roum. Math. Pure. Appl.,* vol. 29, pp. 593–598, 1984.

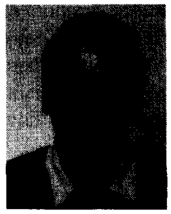
[22] W. E. Schmitendorf, "Cooperative games and vector-valued criteria problems," *IEEE Trans. Automat. Contr.,* vol. AC-18, Apr. 1983.

[23] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica,* vol. 33, pp. 520–534, 1965.

[24] W. Zangwill, *Nonlinear Programming.* New York: Wiley, 1969.

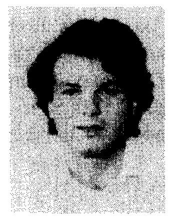
Ravi Mazumdar (S'80–M'88) was born in Bangalore, India. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, India, in 1977, the M.Sc., DIC in control systems from Imperial College, London, England, in 1978, and the Ph.D. degree in systems science from the University of California, Los Angeles, in 1983.

From 1978 to 1979 he was employed with GEC Electrical Projects Ltd., Rugby, England. From April 1983 to Oct. 1983 he was a Member of Technical Staff, AT&T Bell Laboratories, Holmdel, N.J. He held visiting appointments in the Department of Systems Science, UCLA in 1983–1984 and with the Department of Applied Mathematics, Twente University, Enschede, The Netherlands, in the Fall of 1984. From 1985–1988 he was an Assistant Professor in the Department of Electrical Engineering, Columbia University, New York, NY, where he was a member of the Center for Telecommunications Research. In July 1988 he joined INRS-Telecommunications, Montreal, P.Q., Canada, where he is currently an Associate Professor. He is also an Auxiliary Professor in the Department of Electrical Engineering, McGill University, Montreal. His research interests are in the modeling and analysis of telecommunication systems, stochastic filtering and control, and applied stochastic processes.



Lorne G. Mason was born in 1942 in Saskatchewan, Canada. He received the B.Sc. and Ph.D. degrees in mechanical engineering from the University of Saskatchewan in 1963 and 1972, respectively.

He was with Bristol Aerojet, Winnipeg, Man., Canada, from 1963 to 1965 involved in the design of the Black Brant rockets. He joined the British Columbia Telephone Co. as a Traffic Engineer in 1966 and again in 1972 as a consultant for planning digital networks. In 1973 he served as a consultant to Yale University, where he, in collaboration with Professor K. S. Narendra, pioneered the use of learning automata for adaptive routing in telecommunication networks. Between 1974 and 1977, he was with Bell-Northern Research, where he developed planning tools and methods for digital network evolution and state dependent routing. Since 1977 he has been with INRS-Telecommunications, University of Québec, P.Q., Canada, where he currently holds the title of Full Professor. His primary research interests are in the application of control theory and operations research methods to telecommunication network design and management. He has held numerous industrial research contracts and NSERC strategic grants in the area of broadband network design and analysis and has numerous publications on the subject. He is a participant in the recently announced Network of Centers of Excellence program sponsored by the Canadian Government, where he is project leader on broad-band network control.



Christos Douligeris was born in Arcadia, Greece, in 1962. He received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1984, and the M.S., M.Phil., and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, in 1985, 1987, and 1990, respectively.

Since August 1989 he has been with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, as an Assistant Professor. His research interests include modeling and control of telecommunication networks, performance evaluation of queueing systems, and game theory.

Dr. Douligeris is a member of ORSA and the Technical Chamber of Greece.