PRELIMINARY DRAFT
NOT FOR CIRCULATION

ON THE STABILITY OF EFFICIENT NETWORKS:

Integration and Fragmentation in Communication and Transportation Networks*

Glenn A. Woroch

Economics & Statistics Department
GTE LABORATORIES INCORPORATED
Waltham, MA 02254

July 1990

Abstract

Stability of efficient networks is studied by applying cooperative game theory to a three-node system. Conditions are sought which ensure the existence of a core to the corresponding welfare game. I depart from existing literature by examining three cases that capture features common to communication and transportation industries: (i) nonseparable economies of network integration, (ii) realistic restrictions on price structure, and (iii) imperfect substitutability among alternate routes. Each feature shifts the strategic balance between the incumbent network and potential entrants. I derive conditions under which entry is profitable.

---

* Presented at ORSA/TIMS Conference, Denver, October 1988 and Fifth World Conference on Transportation Research, Yokohama, Japan, July 1989. The views expressed in this paper are my own and should not be attributed to GTE Corporation nor any of its subsidiaries.

1. INTRODUCTION

Large, integrated networks have inherent advantages over a collection of smaller, interconnected systems. Most importantly, a single entity can exploit cost and demand complementaries in the delivery of multiple services.

The scale economies which pervade network technologies is the principal source of the advantages of integration. Usually, establishing service between two sites exhibits falling unit costs. By eliminating some links, a network operator can collect traffic from several locations, and route it over common facilities.

Cost savings of this sort abound in communications and transportation. In transportation, inbound and outbound traffic ply the same rights of way. Similarly, modern telecommunications technologies allow voice and data transmissions to double up over the same circuits and radio frequencies. In both industries, traffic flows vary over time allowing facilities to serve both peak and off-peak demands. Similarly, long-haul and short-haul traffic can traverse the same path more cheaply than two dedicated lines.

An integrated network can also achieve higher levels of reliability through coordinated planning and operation. Links can fail (e.g., due to a serious traffic accident), and so too can nodes (e.g., an airport closure). The "star" architecture common in telephone networks is particularly vulnerable to switch failures; in comparison, "ring" and "mesh" topologies offer many alternative routing options. Finally, excess capacity that arises due to variable loads can be minimized by reallocating traffic to idle facilities.[1] A single supplier can select the appropriate system-wide architecture and coordinate traffic flows with a minimum of traffic disruptions.

Complementaries also arise on the demand side. Travelers and phone users are likely to view incoming and outgoing traffic as highly complementary. The same may be true for co-terminus segments that they combine to complete longer trips.

Access to a network confers a demand externality as additional nodes expand the feasible origins and destinations of traffic flows.[2] Such spillovers occur when a friend decides to subscribe to telephone service, or when a low-cost co-generator sells back electricity to the power grid. As long as inter-operability is less than perfect, partitioning users among several networks forgoes the full benefit of this externality.

In addition, proliferation of networks multiplies transaction costs. Loading and unloading are unavoidable when freight and passengers must switch suppliers.[3] Signal and protocol conversions are often required to transfer communications traffic from one network to another. Furthermore, discrepancies in suppliers' traffic schedules result in longer delays and possibly aborted connections.

Despite all the advantages of integration, network industries have properties that make them prone to fragmentation. Scale economies along routes will eventually to exhausted as the complexities of handling ever higher volumes grow. In addition, the limited routes offered by sparse networks will force users to compromise. Given wide differences in distance, duration and quality of the service, users will be attracted to specialized networks which accommodate their personal needs.

Entrants can mount a serious threat to an integrated network by exploiting product-specific scale economies for a limited set of services. Market penetration is made easier by the integrated network's sparse architecture which offers users less-than-ideal routing options.

New entrants have many market niches from which to choose.[4] Of course, by itself, a large number of potential markets cannot guarantee success. The challenge facing the cost-minimizing network is to seal off these niches from cost-raising entry.

My goal is to examine when efficient networks are stable under free entry. To model the competitive process, I take the approach of cooperative games. In this framework stability reduces to the existence of a "core" of a surplus sharing game. This occurs when surplus can be allocated in such a way that no subset, or "coalition," of users can individually improve their welfare by splitting off and supplying services through a separate network.

Most often, strategic formulations of network equilibrium examine the selection of paths between origin and destination pairs over a fixed network. In a rare application of cooperative games, Kalai and Zemel (1982) find conditions for existence of a strong form of the core when network flows freely combine to form coalitions.

More recently, cooperative games have been used to study selection of network designs and the allocation of common costs among traffic flows. Bird (1976) and Granot and Huberman (1981) establish existence of a core when links can handle an unlimited capacity for a fixed cost. Granot and Hojati (1990) devise a strong sufficient condition for a nonempty core when link costs are proportional to traffic levels. When costs are convex, Sharkey (1988) finds conditions which ensure the network cost function is "supportable." This property reduces to the existence of a core of a cost-sharing game played by flows. Bittlingmayer (1990) verifies core existence for a 3-node network in which unit link costs are falling and the symmetric flow demands are price elastic.

These studies examine fairly general network topologies and service offerings. They make several restrictive assumptions on user tastes and network technology, however. In contrast I study a simple 3-node network but allow for more realistic demand and cost conditions. This topology is rich enough to capture many network phenomenon (e.g., alternate routing, and integrated planning and operation).[5] Along the way I point out how these features correspond to communications and transportation networks.

First I drop the assumption of separability of costs for links and nodes. By assuming that capital outlays and operating expenses are additively separable across links, the models of network stability ignore important cost synergies that arise from a contiguous system.

Next I consider pricing restrictions that better approximate realistic rate structures. For instance, networks often lack information needed to discriminate among users; other times they are barred from doing so by regulation.

Finally I allow users to substitute between different routes. When the extra time, distance and inconvenience of indirect routes make them inferior to direct routes, users will migrate to specialized carriers which offer them better service.

Each of these three features alters the strategic balance between incumbent network and potential entrants. In some cases they hinder the incumbent's ability to deflect an entrant, while in some cases it opens an arbitrage possibility that cannot be closed. In each case I derive conditions on demand and cost which indicate when entry becomes profitable.

2. A SIMPLE NETWORK

2.1 Traffic Flows

Consider the network depicted in Figure 1. Traffic originates at each of three nodes destined for both of the two remaining nodes. Index nodes by $i = 1,2,3$ and let the ordered pairs of nodes (i,j) denote the six origin-destination (O-D) pairs. A total of twelve directed, non-cyclic paths exist: six direct $\langle i,j \rangle$ and six indirect $\langle i,k,j \rangle$ where $k \neq i,j$. Examples of both are pictured in Figure 1.

The reason for the notational distinction between the O-D pairs and the paths becomes clear when we describe traffic flows. Let x_{ij} be the traffic requirement for O-D pair (i,j). A portion of this traffic could take an indirect route $\langle i,k,j \rangle$ through the remaining node k; denote this amount by y_{ij} where $0 \leq y_{ij} \leq x_{ij}$. Then the traffic carried on (directed) link $\langle i,j \rangle$ includes direct traffic plus all indirect traffic passing through from elsewhere in the network. When there are three nodes, this is unambiguously expressed as:

$$(1) \qquad z_{ij} = x_{ij} - y_{ij} + y_{ik} + y_{kj}$$

where $k \neq i,j$. Total traffic volume in both directions between i and j is given by:

$$(2) \qquad \bar{z}_{ij} = z_{ij} + z_{ji} = \bar{x}_{ij} - \bar{y}_{ij} + \bar{y}_{ik} + \bar{y}_{kj}$$

where $\bar{x}_{ij} = x_{ij} + x_{ji}$ and $\bar{y}_{ij} = y_{ij} + y_{ji}$ are the bi-directional levels of direct and indirect traffic.

I assume that the span of time is short enough so that the level of traffic flows on any given O-D pair is fixed. The problem of synthesis arises when a network of fixed capacity must be installed to accommodate varying traffic levels over time.[6]

2.2 Cost of Service

In general, the cost of service depends on a complete description of traffic flows. Given the flows, the network operator selects an architecture, builds the corresponding links and nodes, and then routes the traffic over available paths. Rather than having users non-cooperatively choose the cost-minimizing path, the network operator dictates a routing scheme and users select the carrier offering the best price-service package.

Traffic from as many as six O-D pairs travel between any two nodes so that cost depends on vectors $x = (x_{ij})$ and $y = (y_{ij})$. Without loss of generality, cost can be decomposed into fixed and variable portions, $F(A) + V(x,y)$, where the network architecture is given by $A \equiv \{(i,j): z_{ij} > 0\}$. When traffic direction matters, any one of $2^6 - 1 = 63$ architectures is feasible. If cost depends solely on the sum of traffic in both directions, there are just $2^3 - 1 = 7$ architectures, three each for one-link and two-link networks plus the complete three-link system.

The fixed cost $F(A)$ includes expenditures required to construct nodes and links. Included are expenses to acquire access to public thoroughfares and to prepare the rights of way. These costs do not vary with the traffic level, but differences in terrain and climate could lead to differences for travel over comparable distance.

I assume monotonicity of fixed costs:

$$(3) \qquad S \subset T \text{ implies } F(S) \le F(T)$$

This is a weak assumption which says that incremental cost of an additional link (and the accompanying nodes) is always positive. This is highly plausible since a smaller network could be built by simply halting construction of a larger network short of completion.

A stronger condition is <u>subadditivity</u>. In that case, savings are available to building a large network rather than with two separate ones. Subadditivity of F is equivalent to assuming <u>construction</u> of the network is a natural monopoly; it does does not guarantee natural monopoly in the provision of network services.

Variable cost $V(x,y)$ includes operating and maintenance expenses, and also volume-sensitive construction outlays. Assume that V everywhere nondecreasing in its arguments.

The network operator's cost-minimizing construction and routing plan solves:

$$(4) \qquad C(x, y) = \min \left\{F(A) + V(x,y) \colon 0 \le y_{ij} \le x_{ij}\right\}$$

The optimal routing pattern y^*_{ij} implies an optimal architecture A^* which will, in general, differ from traffic requirements $A^+ = \{(i,j)\colon x_{ij} > 0\}$. The two sets may even be disjoint: for instance, the cheapest way of handling the traffic x_{13} could be routing it along the path $\langle 1,2,3\rangle$ using the architecture $A^* = \{(1,2),(2,3)\}$. If all costs are fixed, this happens if $F(\{(1,3)\})$ is much larger than $F(\{(1,2),(2,3)\})$. Similarly, the complete network could be efficient in spite of its higher fixed costs due to rising variable costs on heavily traveled links.

Given a pattern of direct and indirect flows, solving (4) is a simple matter of building the necessary network to handle the requirements. No solution exists to the more general problem of simultaneously designing and routing the traffic.[7] It is the nontrivial solution to the routing problem, coupled with the fixed costs of connection, that distinguishes our framework as a network analysis.

In real networks, costs can be highly dependent on apportionment of traffic by direction. For instance, pumping oil uphill is far most costly than letting gravity take its course. Also cable television systems usually adopt uni-directional technologies. Direction is less important for telephone and electricity networks. When it does not matter, we may write costs as $F(A) + V(\bar{x},\bar{y})$ where $A = \{(i,j): \bar{z}_{ij} > 0\}$. A simple derivation establishes that, when fixed and variable costs depend on traffic aggregated by direction, so too will the network cost function, i.e., $C(x,y) = C(\bar{x},\bar{y})$. In addition, the optimal routing pattern will have traffic between two nodes all travel direct routes, or all indirect, but not a mixture, i.e., $\bar{y}^*_{ij} = 0$ or $\bar{y}^*_{ij} = \bar{x}_{ij}$.

2.3 Service Demands

Users derive benefits from total flow between O-D pairs. Actually, each node could be a gateway for many users who collectively value connection with the other two nodes. They also may value traffic between two distant nodes, if only as an indirect route to their final destination.

Index the finite number of users by $n \in N$. The benefit they derive from flows is given by $B^n(x^n,y^n)$ where $x^n = (x^n_{ij})$ and $y^n = (y^n_{ij})$ are vectors of direct and indirect flows for user n. This formulation can describe many situations. To express users' (exclusive) preference for round trips, write $B^n(x^n_{ij},x^n_{ji})$. Imperfect substitution between direct and indirect routes is captured by $B^n(x^n_{ij},y^n_{ij})$.[8]

Several restrictions underlie this specification. First of all, I implicitly assume that users' locations are fixed in advance of network construction. If they locate afterwards, then aggregate demands would be a function of the architecture selected by the network operator.

Second, traffic has been treated as essentially a one-dimensional quantity. In fact shipments consist of a wide variety of goods differing in density and perishability, and endless voice and data messages wind their way through the telephone network.

Finally, income effects have been ignored. When lump-sum transfers are possible this will allow us to re-distribute wealth across users.

3. THE COOPERATIVE GAME APPROACH

Fundamental to the cooperative-game approach is the notion of a coalition, a subset of the full set of players: $S \subset N$. In practice, coalitions can form to compete with the integrated network in several ways. Users may ban together to realize reduced cost from a smaller network. Alternatively, perceiving an arbitrage possibility, a third party may build a network and signup subscribers in competition with the public network. Either way, the new supplier has access to exactly the same technology as the incumbent.

To assess the payoff to a stand-alone network, we need to measure the surplus available to each coalition. Since there are no income effects, and supposing that lump-sum transfers of wealth across users are feasible, we can express benefits to a coalition in the reduced form:

$$(5) \qquad \sum_{n\in S} B^n(x^n,y^n)$$

for each $S \subseteq N$. Notice that in benefit terms, traffic of different individuals is additive.

Adopting the notion of a welfare game from Sharkey (1982a), the characteristic function assigns to each coalition the maximum consumer surplus that it can achieve:

$$(6) \qquad W(S) = \max_{(x^n),(y^n)} \{ \sum_{n\in S} B^n(x^n,y^n) - C(x^S,y^S)\}$$

where $x^S_{ij} = \sum_{n\in S} x^n_{ij}$. Since null flows (i.e., $x^n_{ij} = 0$ for all i,j) are always feasible, we have $W(S) \geq 0$. Indeed, involuntary dropoff may be part of the solution to this problem if the net cost to some user's participation exceeds his net benefit to the coalition. Since flows are managed, these users are simply denied access to the network.

The <u>core</u> of the welfare game (W,N) consists of all surplus vectors (w^n) satisfying :

$$\sum_{n \in N} w^n = W(N) \tag{7}$$

$$\sum_{n \in S} w^n \geq W(S) \quad \text{for all } S \subset N \tag{8}$$

Embodied in this definition of stability is a notion of competition by stand-alone firms:

(i) No modal split: users cannot purchase a portion of their service requirement from one network and the remainder from another. In practice shippers often use different carriers along the same route; phone users may subscribe to more than one long-distance service. Recalling the distinction between architectures and coalitions, a successful entrant may not have any users in common with an incumbent even though their networks overlap along certain routes.

(ii) No supplier switching: users cannot combine partial segments from the two networks to form a complete trip. In contrast, shippers regularly secure end-to-end rail service through inter-lining agreements of two or more railroads. Similarly, long-distance calls are regularly transmitted over the local networks at the originating and terminating ends with an interexchange carrier in between.

(iii) No network interaction: traffic on one network has no affect on the cost or service of the competing network even if their architectures overlap. This would not hold when several air carriers route passengers through congested airport facilities.

The first task is to describe conditions on demand and cost which ensure that a core will exist. There are very general results which underscore how demand and cost complementarities promote stability. For instance, Sharkey (1982a) shows how if each B^n displays complementarity across the x^n_{ij} and y^n_{ij}, and the cost function C exhibits cost complementarity in x^N_{ij} and y^N_{ij}, then the core of the game (W,N) is nonempty.

These sufficient conditions are often difficult to verify; other times they are plainly unrealistic. An alternative approach gives necessary and sufficient conditions for core existence based on the "balancedness" of the characteristic function (Shapley (1967)). The drawback of this approach, as we will see, is that it requires calculation of the characteristic function. Nevertheless these conditions can be extremely powerful. For example, when there are three players, provided that the function W is superadditive, a necessary and sufficient condition for the core to exist is:

$$(9) \qquad W(\{1,2\}) + W(\{1,3\}) + W(\{2,3\}) \leq 2W(\{1,2,3\})$$

where 1,2,3 denote the players (Moulin (1982)).

To exploit this condition, I will often consider a special case which reduces to a game of three players. Imagine that two types of users reside at each node, one desiring to travel to each of the remaining two nodes. Assume that users traveling between two nodes are identical so that they can be treated as a single entity.

Furthermore, traffic flows in many networks have a great degree of directional symmetry. Round trips in passenger transportation and two-way voice communication ensure that incoming and outgoing flows are nearly equal.

This makes three groups. Ignoring traffic direction, label links (1,2), (2,3), and (1,3) as 1, 2, and 3, respectively. This number scheme will denote both user groups and links.

4. INTEGRATION ECONOMIES

I begin to explore network stability by introducing integration economies. I retain scale economies along individual links, but unlike other cooperative game models, the cost of service need not be additively separable across links.

To focus on cost factors, we want to eliminate benefits from the comparison in (9). This can be done in several ways. First, when there are no transmission costs, a coalition maximizes its wealth by equating the marginal benefit of each user to zero: $B_i'(\hat{x}_i) = 0$. The same happens when demand is perfectly price inelastic, in which case every user enjoys his maximum traffic $x_i = \hat{x}_i$. Either way, the value of the benefits portion will be the same on both sides of the inequality (9).

4.1 Subadditive Fixed Cost

Before addressing the stability issue, let us probe the source of integration economies a little further. In the absence of transmission costs, integration economies exist when a single network can provide the services of two or more smaller networks at lower cost. Mere subadditivity of fixed costs of constructing an architecture is not entirely enough for this.

The fixed cost function F(A) introduced earlier is the solution to a minimization problem. It finds the least cost of constructing a collection of links and nodes to complete the desired architecture. Formally, it solves:

$$F(A) = \min \{\sum_{\alpha} f(A_\alpha) : \bigcup_{\alpha} A_\alpha \supseteq A\}$$

The sets of links and nodes A_α are themselves architectures which, when joined together, provide the same end-to-end services as A. I require only that the cost of constructing each of these subnetworks, $f(A_\alpha)$, is monotone in the A_α. It is a simple matter to prove that F not only inherits monotonicity from f, but that it is subadditive as well:

(10) $$F(A) + F(A') \geq F(A \cup A')$$

where $A \cap A' \neq \phi$. The formal argument is spelled out in the Appendix. Intuitively, the minimum cost should be no more than the sum of disjoint subnetworks A_α that partition architecture A. Of course this requires that the subnetworks can be interconnected at no extra cost.

Such economies arise when equipment at network nodes perform multiple functions for a small amount more than specialized equipment. In communications networks, computers perform switching, signalling and database functions. Loading and offloading, storage, and transfer occur at terminal facilities in railroad networks.

In the simple three-user network, the network cost function inherits subadditivity from F(A). Without loss of generality, assume:

(11) $$F_{12} < F_{23} < F_{13}$$

renumbering links if necessary. Inequalities (10) and (11) then imply that F_1, $F_2 \leq F_{12} \leq F_{123}$ so that the network $A^* = \{1,2\}$ is the cost-efficient architecture for the grand coalition. The same is true for two-user coalitions. For i = 1 or 2, $C(\{i\}) = F_i$ since $F_i < F_{12} \leq F_{ij}$. Necessarily, $C(\{1\}) + C(\{2\}) = F_1 + F_2 \leq F_{12} = C(\{\{1,2\})$. We can only claim that $C(\{3\}) = \min\{F_3, F_{12}\}$. Nevertheless $C(\{i\} + C(\{3\} = F_i + \min\{F_3, F_{12}\}$ $\leq \min\{F_i + F_3, F_{12}\} = C(\{i,3\})$. Thus, in the absence of transmission costs, the network cost function is subadditive.

In addition to subadditivity, the network cost function is stable as well: the left side of (9) equals $3F_{12}$ and the right side $2F_{12}$. Clearly the ability of different users to share the same architecture without any additional cost is a strong inducement to join together. This logic does not follow when diverted traffic incurs transmission costs, however.

4.2 **Constant Marginal Transmission Costs**

At this point I introduce transmission costs in a simple way through linear terms: $V(x,y) = \sum_i v_i \bar{z}_i$ where $\bar{z}_i$ is given by (2). Then the network cost function becomes:

$$
\begin{aligned}
C(x^{\{i\}}) &= \min\{F_i + v_i x_i,\ F_{jk} + (v_j + v_k)x_i\} \\
C(x^{\{i,j\}}) &= \min\{F_{ij} + v_i x_i + v_j x_j,\ F_{ik} + v_i x_i + (v_i + v_k)x_j, \\
&\qquad F_{kj} + (v_k + v_j)x_i + v_j x_j\} \qquad (12) \\
C(x^N) &= \min\{F_{123} + \sum_{\ell=1}^{3} v_\ell x_\ell,\ F_{ij} + v_i x_i + v_j x_j + (v_i + v_j)x_k\}
\end{aligned}
$$

The complete network is dominated in the case of a two-user coalition because one link will go unused but all other architecture and routing alternatives are possible. Without deriving an explicit form, we can characterize the solutions to the cost minimization problems as:

$$C(x^{\{i,j\}}) = F^*(\{i,j\}) + \sum_{\ell=i,j} v_\ell^*(\{i,j\})x_\ell$$

where:

$$F^*(\{i,j\}) = \begin{cases} F_{ij} & x_i \text{ and } x_j \text{ both direct} \\ F_{jk} & x_i \text{ indirect, } x_j \text{ direct} \\ F_{ik} & x_j \text{ indirect, } x_i \text{ direct} \end{cases}$$

$$v_k^*(\{i,j\}) = \begin{cases} v_i + v_j & x_k \text{ indirect} \\ v_k & x_k \text{ direct} \end{cases}$$

The fundamental tradeoff when selecting an alternative is a comparison of the incremental construction cost of an indirect architecture against the incremental transmission cost from roundabout routing. For instance, with three users, the complete network may be cheapest if:

$$F_{123} + \sum_{\ell=1}^{3} v_\ell x_\ell \; < \; F_{ij} + v_i x_i + v_j x_j + (v_i + v_j) x_k$$

A rearrangement reveals the tradeoff between fixed and variable costs:

$$F_{123} - F_{ij} < (v_i + v_j - v_k) x_k$$

The complete network could dominate <u>all</u> combinations (i,j,k) so long as the x_k's are large enough and the $(v_i + v_j - v_k)$'s are all positive. In other words, unit transmission costs must be close to one another.

Tedious calculations in the Appendix verify that the network cost function $C(x^S)$ is subadditive, so that a single, integrated network is the least-cost industry structure. However, as is well know (Faulhaber (1975)), the strong complementarities which create a natural monopoly will not always protect it against entry.

What additional conditions are needed to ensure the efficient network is stable? We return to condition (9) for core existence with three users. Straightforward computation shows that $\sum_{i<j} C(\{i,j\}) = \sum_\ell v_\ell^* x_\ell$ where $v_\ell^* \in \{2v_\ell, \sum_m v_m, v_m + v_n\}$, reducing the number of possibilities for the left side of (9).

To pare down the possibilities further, I consider the case of symmetric demands: $x_i = \hat{x}$ for all i. Then the integrated network is stable provided two conditions hold:

(13) $$2F_{123} < \sum_{i<j} F_{ij}$$

(14) $$F_{ik} + F_{kj} - F_{ij} < 2(v_k - v_i - v_j)\hat{x} \quad \text{for all } i, j, k$$

The second inequality makes clear the role of transmission costs. The first does so as well since it implies the fixed cost of direct service is not too large.

5. PRICING RESTRICTIONS

Above, I implicitly relied on user-specific, lump-sum charges to sustain the efficient configuration. These charges effect inter-personal transfers giving the integrated network great freedom to attract and retain users. A scheme of this sort requires global information about each user's willingness to pay. Even when users having the same origin and destination can be aggregated, charges for each of twelve directional flows must be devised.

Often network operators do not have the ability to segregate traffic by route. Shipments of electricity and natural gas become inextricably mixed as they travel through their networks. Passengers and freight can become difficult to trace when partial trips are purchased from one or more carriers. This is especially true when re-sale markets for trips segments become active. Furthermore, prices may not discriminate among traffic flows on a given link, as when non-stop passengers share the same plane with those passing through a hub.

Prices are further restricted by regulatory mandate of nondiscrimination as a means of achieving goals of equitable income distribution. This concern has surfaced recently when policymakers have expressed fear that airlines operating hub-and-spoke networks overcharge traffic that originates or terminates at a hub.

Certainly these pricing restrictions limit the ability of the integrated network to hold onto customers through surplus redistribution.

The entrant's ability to profit from a rate distortions is hampered if it is likewise constrained. I find that nondiscriminatory prices lends stability to the efficient network in some cases but not others.

I assume demands are inelastic and there are no transmission costs. Let r_i represent the price paid by user i for service x_i on route i. This formulation fails to capture the full extent of network pricing restrictions--such as usage sensitive tariffs--but it illustrates the effect on strategies of the incumbent and entrant alike.

To begin, single-user coalitions are unaffected by nondiscriminatory rates. Coalitions with two users experience a loss of welfare when some traffic is indirectly routed. In the coalition $S = \{i,j\}$, users receive $b_i - r_i$ and $b_j - r_j$, respectively, where $r_i + r_j \geq F_{ij}$ to break even. The surplus $b_i + b_j - F_{ij}$ can be distributed in any way between the two users provided $r_j \geq 0$.

Now suppose that user i receives indirect service and user j direct. Their surpluses are $b_i - r_j - r_k$ and $b_j - r_j$, respectively, subject to the budget constraint $2r_j + r_k \geq F_{jk}$. Notice that they pay the same amount for use of the shared link. As a result, indirect traffic always contributes at least as much as direct traffic to defraying the common costs.

This restriction eliminates certain surplus distributions that were available under route-specific pricing. This is not evident for the coalition $S = \{1,2\}$ since it adopts the least-cost architecture. Consider instead the coalition $S = \{2,3\}$. By routing user 3 indirectly, the coalition can achieve the same construction costs, F_{12}. Since user 3 must pay at least as much as user 2, assignment of this cost is constrained. In the extreme, they share the cost equally so that both earn a surplus of b_i

- $F_{12}/2$. Higher surplus for user 3 is possible, but only by selecting architecture {2,3}. Upon routing all traffic directly, the burden can be shifted away from user 3 at an incremental cost of $F_{23} - F_{12}$. Surplus possibilities are illustrated in Figure 2.

Limitations on surplus redistribution are even more dramatic for the grand coalition. Once again the full range of allocations is possible when direct service is provided to each user, albeit at the high cost of F_{123}.

Suppose instead that architecture {2,3} is selected, so that user 1 is indirectly routed. Payoffs are then $w_1 = b_1 - r_2 - r_3$, $w_2 = b_2 - r_2$, and $w_3 = b_3 - r_3$ where $2(r_2 + r_3) \geq F_{23}$. Charges should never exceed joint cost so that the budget constraint necessarily binds. In that event $w_1 = b_1 - F_{23}/2$ regardless of the rates chosen. Provided all three users are served, the only degree of freedom left to the network operator is how to distribute the remaining cost of $F_{23}/2$ between users 2 and 3.

To assess the stability of the integrated network, we trace out the frontier of surplus possibilities for the grand coalition and check whether one- or two-user coalitions can improve their lot by departing.

Even though it is more costly, direct service could be efficient if the pricing restrictions are sufficiently tight. One-user coalitions dominate when $r_i \leq F_i$, $i=1,2$ and $r_3 \leq \min(F_3, F_{12})$. By subadditivity, $\sum_i r_i = F_{123} \leq \sum_i F_i$, so that some surplus allocations can fend off one-user coalitions as long as $F_3 \leq F_{12}$. Otherwise stability is problematic.

Turning to two-user coalitions, we find that every member will opt for the {1,2} architecture which provides surplus $b_i + b_j - F_{12}$ for at least some tariff charged to members of the defecting coalition. Therefore, a stable integrated network must have $r_i + r_j < F_{12}$ for each pair of users i and j. Thus $2(r_1 + r_2 + r_3) < 3F_{12}$. Substituting the breakeven constraint, $\sum r_i = F_{123}$, we have the necessary condition for stability: $F_{123} < 3F_{12}/2$.

We must also examine stability when the integrated network opts for one of the two-link architectures. The most likely candidate for an efficient configuration is once again the {1,2} architecture. Recall that, in this case, user 3 receives a fixed allocation of surplus $w_3 = b_3 - F_{12}/2$. This arrangement will be upset by defection of a single user whenever $F_3 < F_{12}/2$, a possibility that cannot be ruled out.

Suppose, instead that the stand-alone cost of direct service for user 3 is too high to justify his departure. Then users 1 and 2 share the remaining cost: $w_1 + w_2 = b_1 + b_2 - F_{12}/2$. Since the best that a two-user coalition can achieve has its members sharing twice this amount, $w_1 + w_2 = b_1 + b_2 - F_{12}$, defection is never worthwhile.

The other two-link architectures are not so easy to defend. First of all they must also fend off one-link attacks which requires $F_i < F_{jk}/2$. As before, indirect traffic (e.g., user i) bears half the fixed cost while direct traffic (e.g., users j and k) shares the remainder. Two-user coalitions will again build a {1,2} architecture. A simple calculation shows that challengers can charge user ℓ as little as $r_\ell = F_{12} - F_{jk}/2$. Therefore, a necessary condition for stability is that the integrated network makes them a better offer is: $F_{jk}/2 < 2F_{12} - F_{jk}$. Equivalently, $3F_{jk}/4 < F_{12}$ for all j,k.[9]

6. IMPERFECT ROUTE SUBSTITUTION

Up to this point, users were concerned only about total flow between some origin and destination, and not with the route taken. This assumption is often unrealistic. First of all indirect routes are usually more time consuming. The extra time raises trip cost for passengers and can lead to spoilage for perishable freight. Circuitous routes also cover a greater distance which can lead to higher line loss for gas and electricity transmissions and increased breakage for fragile shipments. Finally, the multiple connections reduce the overall service quality of the trip. Certainly this is well known to air travelers who have made enroute plane changes. Also, repeated loading and unloading of delicate items will undoubtably raise damage rates.[10]

A penalty for inferior routing can be incorporated in a simple way. Once again, let demand be perfectly price inelastic, but now assume that direct service provides a total benefit of b_i to user i; indirect service (via links j and k) confers benefit b_i' where $b_i' < b_i$.

This breaks the identity between a traffic flow and a player in the cooperative game. In this case the efficient network is the one that provides the greatest total surplus of benefits over costs. The characteristic function can now be expressed as follows:

$$W(\{i\}) = \max \{b_i - F_i,\ b_i' - F_{jk},\ 0\}$$

$$W(\{i,j\}) = \max \{b_i + b_j - F_{jk},\ b_i + b_j' - F_{ik},\ b_i' + b_j' - F_{123},\ W(\{i\}),\ W(\{j\}),\ 0\}$$

$$W(\{1,2,3\}) = \max \{\sum_i b_i - F_{123},\ b_i' + b_j + b_k - F_{jk},\ b_i' + b_j' + b_k - F_{123},$$
$$\sum_i b_i' - F_{123},\ W(S),\ 0\}$$

Recalling the convention that $F_{12} < F_{23} < F_{13}$, and using the assumption $b_i' < b_i$, many elements in the maximization problems can be eliminated:

$$W(\{i\}) = \max \{b_i - F_i,\ 0\}$$

$$W(\{3\}) = \max \{b_3 - F_3,\ b_3' - F_{12},\ 0\}$$

$$W(\{1,2\}) = \max \{b_1 + b_2 - F_{12},\ b_1 - F_1,\ b_2 - F_2,\ 0\}$$

$$W(\{2,3\}) = \max \{b_2 + b_3 - F_{23},\ b_2 + b_3' - F_{12},\ b_2 - F_2,\ b_3 - F_3,\ 0\}$$

$$W(\{1,3\}) = \max \{b_1 + b_3 - F_{13},\ b_1 + b_3' - F_{12},\ b_1' + b_3 - F_{23},\ b_1 - F_1,\ b_3 - F_3,\ 0\}$$

$$W(\{1,2,3\}) = \max \{\ \sum_i b_i - F_{123},\ b_i' + b_j + b_k - F_{jk},\ b_i - F_i,\ 0\}$$

Note that indirectly routing two or more flows will always reduce surplus since indirect traffic forgoes some utility, and yet the complete network must be built.

By inspecting these expressions, we can verify that the characteristic function is superadditive. This allows us to apply (9) to check whether the resulting cooperative game has a nonempty core. Given assumptions on the F's, the least-cost architecture consists of links 1 and 2. Although all users are served, user 3 suffers a surplus loss equal to $b_3 - b_3'$ compared with that offered by the complete network.

Note that if the grand coalition ever finds it worthwhile to serve more than one user group, then it pays to serve all three. Also, the complete network will be built only if all flows are routed directly: two or more indirect routes lower surplus without any cost reduction.

To find conditions that characterize the core, all possible values for both sides of (9) must be considered.

Aside from one-flow networks, the grand coalition has four possible configurations for the efficient network. Before examining the implications of each one in turn, a few plausible assumptions will greatly reduce the number of comparisons that are necessary. Let:

(15) $$b_i < F_{ij} - F_i$$

(16) $$b_i' < F_{kj} - F_j$$

These two conditions state that the incremental cost of extending direct (indirect) service to user i exceeds the additional cost of already serving user j. As a result, a stand-alone network--whether it provides direct or indirect service--is justified by the benefits it offers.

Case 1: First suppose that the welfare loss to user 3 cannot offset the cost savings from the least-cost network: $W(\{1,2,3\}) = b_1 + b_2 + b_3' - F_{12}$. By comparing this configuration with the other three alternatives, we see that it achieves a higher level of welfare when, among other conditions, $b_3 - b_3'$ $< F_{23} - F_{12} + (b_1 - b_1')$. This inequality rules out one of the three possibilities in the maximization problem defining $W(\{1,3\})$, leaving:

$$W(\{1,3\}) = \max \{b_1 + b_3 - F_{13},\ b_1 + b_3' - F_{12}\}$$

From above we know that:

$$W(\{1,2\}) = b_1 + b_2 - F_{12}$$

$$W(\{2,3\}) = \max \{b_2 + b_3 - F_{23},\ b_2 + b_3' - F_{12}\}$$

The left side of (9) will take one of four values:

$$2(b_1 + b_2 + b_3) - \sum_{i<j} F_{ij}$$

$$2(b_1 + b_2) + b_3 + b_3' - 2F_{12} + F_{13}$$

$$2(b_1 + b_2) + b_3 + b_3' - 2F_{12} + F_{23}$$

$$2(b_1 + b_2 + b_3') - 3F_{12}$$

When compared with the right side of (9), at least one of these conditions holds when:

$$(17) \qquad b_3 - b_3' < \frac{1}{2}(F_{13} + F_{23} - F_{12})$$

Thus, the core exists when the loss in benefits from diverting user 3 along the indirect route is not too great relative to a certain average of incremental fixed cost over direct routing.

Case 2: Suppose now that the efficient configuration calls for direct routing of all traffic: $W(\{1,2,3\}) = \sum_i b_i - F_{123}$. A necessary condition is $b_i - b_i' > F_{123} - F_{jk}$. In words, the incremental benefit from switching from indirect to direct routing exceeds the incremental cost for each of the users. It is a simple matter to show that these conditions rule out any indirect routing for coalitions of two users. This is expressed as $W(\{i,j\}) = b_i + b_j - F_{ij}$ for all $i \neq j$. Consequently, the condition (9) holds when $\sum F_{ij} \geq 2F_{12}$, or $F_{23} + F_{13} > F_{12}$ which necessarily holds by (11). Unlike case 1, stability of the efficient network is ensured.

Case 3: The last two cases have the remaining users taking indirect routes. First consider the case where user 1 is indirectly routed: $W(\{1,2,3\}) = b_1' + b_2 + b_3 - F_{23}$; then group 3 must be routed directly as well. Consequently, $W(\{i,j\}) = b_i + b_j - F_{ij}$ for $i,j = 1,2$ and $W(\{1,3\}) = \max\{b_1 + b_3 - F_{13}, b_1' + b_3 - F_{23}\}$. Performing the summation on the left side of (9), and comparing with the right side, we find that two conditions must hold for a nonempty core:

(18) $$F_{12} + F_{13} > F_{23}$$

(19) $$b_1 - b_1' < F_{12}$$

The first inequality holds by (11) and the second requires that the incremental benefit to group 1 of direct over indirect routing is not too large.

Case 4: Finally, when $W(\{1,2,3\}) = b_1 + b_2' + b_3 - F_{13}$, we find all routing is direct for two-user coalitions, i.e., $W(\{i,j\}) = b_i + b_j - F_{ij}$. Proceeding as with the other cases, we find that (9) holds when:

(20) $$b_2 - b_2' < \frac{1}{2}(F_{12} + F_{23} - F_{13})$$

A necessary condition for this to hold is:

(21) $$F_{12} + F_{23} > F_{13}$$

This last condition has some bite since it is stronger than (11).

7. CONCLUDING REMARKS

Our findings supply new evidence for a potential conflict between network efficiency and entry incentives. Sparse architectures reduce fixed costs of connecting nodes but raise the variable costs of transporting the traffic along with the user costs from inferior routes and distorted rates. A profit opportunity is left open for a specialized carrier to enter with a limited network. Market constraints on its ability to transfer earnings from elsewhere in the network may leave a full-service provider powerless to stop the incursion.

A principal motivation for this paper was to understand the obstacles to efficiency under free entry taking account of the peculiarities of network industries. This is a crucial public policy issue as deregulation of many of these industries progresses. While the merits of competition propelled the deregulation movement, considerable fear remains that free entry will lead to unnecessary duplication of facilities.

Nevertheless, despite the possibility of de-stabilizing entry and its social cost, policymakers should not immediately resort to government-imposed barriers. Before seeking protection from competition, users and carriers should explore other avenues which might sustain an efficient network.

First of all, the blame may be traced to existing or past regulations. Pricing distortions that invite cost-raising entry are the legacy of years of regulation. Frequently, these regulations bind the hands of incumbents more tightly than entrants, creating profit opportunities where none existed before.

Management of the established networks also shares the responsibility to move the industry towards efficiency. They should explore innovative rate structures, service options, and network technologies to retain existing customers and attract new ones. They must abandon pricing and investment practices adopted under regulatory protection that are ill-suited to a competitive environment.

APPENDIX

A. ***Subadditivity of F(A):*** Let (A''_α) cover $A \cup A'$ where A and A' are disjoint sets of nodes and links. The inequality in (10) is immediate if the A''_αs are themselves mutually disjoint. Suppose, instead, that some pair of A''_α, say A''_1 and A''_2, intersect. Then $A''_1 \cup A''_2 = A''_1 \cup (A''_2 - A''_1)$, and using monotonicity, $f(A''_1) + f(A''_2) \geq f(A''_1) + f(A''_2 - A''_1)$. Also, if for some A''_1, $A''_1 - A \neq \phi$, then cost is lowered by replacing A''_1 with $A''_1 \cap A$. Thus, we can at least assume that efficient configuration of subnetworks A''_αs exactly partition $A \cup A'$. The same applies to (A_α) and (A'_α) that cover A and A' separately. Together the collections (A'_α) and (A'_α) partition $A \cup A'$ as well, but (A''_α) will have lower cost:

$$F(A) + F(A') = \sum_\alpha f(A_\alpha) + \sum_\alpha f(A'_\alpha) \geq \sum_\alpha f(A''_\alpha) = F(A \cup A').$$

B. ***Subadditivity of $C(x^S)$:*** (i) First we must establish that:

$$C(x^{\{i\}}) + C(x^{\{j\}}) \geq C(x^{\{i,j\}}) \tag{A1}$$

for i, j. Each single user will choose either direct or indirect service. If both direct, then the left side of (A1) is $\sum_\ell (F_\ell + v_\ell x_\ell)$ which exceeds $F_{12} + \sum_\ell v_\ell x_\ell$, which in turn exceeds the right side of (A1). If one user chooses direct service and the other indirect, then the left side of (A1) equals $F_j + F_{jk} + (v_j+v_k)x_i + v_j x_j$ which is greater than $F_{jk} + (v_j+v_k)x_i + v_j x_j$, and hence, the right side of (A1).

Finally, users i and j are routed indirectly when $F_{jk} + (v_j + v_k)x_i < F_i + v_i x_i$ and $F_{ik} + (v_i + v_k)x_j < F_j + v_j x_j$, so that $F_i - F_{jk} > (v_j + v_k - v_i)x_i$ and $F_j - F_{ik} > (v_i + v_k - v_j)x_j$, respectively. When $F_i < F_{jk}$ we have a contradiction for these values and so it cannot be efficient to route both users indirectly. The only case not covered has $F_3 < F_{12}$; but then the case is even stronger.

(ii) To finish the exercise we need only verify the inequalities:

$$\text{(A2)} \qquad C(x^{\{i,j\}}) + C(x^{\{k\}}) \geq C(x^N)$$

Again, members of the one-user and two-user coalitions can opt for direct or indirect service; with two routing options for one-user coalitions, and three for two users, there is a total of six possibilities. For most of them, (A2) is straightforward. The following four inequalities treat the case where one of the two networks route all traffic directly:

$$F_{ij} + F_k + \sum_{\ell} v_\ell x_\ell > F_{123} + \sum_{\ell} v_\ell x_\ell$$

$$2F_{ij} + v_i x_i + v_j x_j + (v_i + v_j)x_k > F_{ij} + v_i x_i + v_j x_j + (v_i + v_j)x_k$$

$$F_{ij} + F_k + v_j x_j + v_k x_k + (v_j + v_k)x_i > F_{ij} + v_j x_j + v_k x_k + (v_j + v_k)x_i$$

$$F_{ik} + F_k + v_i x_k + v_k x_k + (v_i + v_k)x_j > F_{ik} + v_i x_i + (v_i + v_k)x_j$$

The left side of the inequalities assume possible values for the left side of (A2); the right sides are at least as large as the right side of (A2). The two remaining cases has some indirect routing in both networks. There, fixed costs are lower using the complete network (by subadditivity) and variables costs are the same.

C. ***Core Existence:*** What matters when forming the left side of (9) is whether each two-user coalition routes all traffic directly, or chooses one of the two indirect architectures. (The two users will never both be indirectly routed, requiring the complete network.) We consider each combination in turn.

(i) Suppose all coalitions choose to route users directly. Then the left side of (9) is simply $\sum_{i<j} F_{ij} + 2(\sum_\ell v_\ell)\hat{x}$. The right side takes on one of two values: $2(F_{123} + \sum_\ell v_\ell \hat{x})$ or $2F_{ij} + 4(v_i + v_j)\hat{x}$. The inequality will hold in the first case if $2F_{123} < \sum F_{ij}$, and in the second if $F_{ik} + F_{kj} - F_{ij} < 2(v_k - v_i - v_j)\hat{x}$.

(ii) If just one of the two coalitions route some traffic indirectly, then Weighting $2F_{123} + 2(v_1 + v_2 + v_3)\hat{x}$ and $2F_{jk} + 4(v_j + v_k)\hat{x}$ equally and summing yields $F_{123} + F_{jk} + (v_i + 3v_j + 3v_k)\hat{x}$ which is smaller than $\sum_{i<j} C(\{i,j\}) = 2F_{jk} + F_{ik} + (v_i + 3v_j + 3v_k)\hat{x}$ since $F_{123} \le F_{ik} + F_{jk}$ by subadditivity.

(iii) When only one of the two-user coalitions route traffic directly, the integrated network is always stable. Three possibilities obtain for the sum of their costs.

$$F_{jk} + 2F_{ik} + (2v_i + 2v_j + 4v_k)\ \hat{x}$$
$$\sum_{i<j} F_{ij} + (2v_i + 4v_j + 2v_k)\ \hat{x}$$
$$3F_{ik} + 4(v_i + v_k)\ \hat{x}$$

Each of these values can be undercut by $2C(\{1,2,3\})$. This is shown by constructing a weighted sum of the possible values for $C(\{1,2,3\})$ given in (12) and showing that it is lower. For instance, the average of $2F_{jk} + 4(v_j + v_k)\hat{x}$ and $2F_{ik} + 4(v_i + v_k)\hat{x}$ is less than the first value above, and hence, inequality (9) must hold. Similarly the average of $2F_{jk} + 4(v_j + v_k)\hat{x}$ and $2F_{jk} + 4(v_j + v_k)\hat{x}$ is less than the second term above. The last term is strictly dominated by $2F_{ik} + 4(v_i + v_k)\hat{x}$.

(iv) If none of the three two-user coalitions route traffic directly, then the sum of their costs takes one of two values:

$$F_{ik} + 2F_{jk} + (v_i + 3v_j + 5v_k)\,\hat{x}$$

$$\sum_{i<j} F_{ij} + 3(v_1 + v_2 + v_3)\,\hat{x}$$

Place a weight of 3/4 on $2F_{jk} + 4(v_j + v_k)\hat{x}$, and 1/4 on $2F_{ik} + 4(v_i + v_k)\hat{x}$; the result dominates the first term above. The second term exceeds $2F_{123} + 2(\sum_i v_i)\hat{x}$ as long as $\sum_{i<j} F_{ij} > 2F_{123}$, the first condition we derived for core nonexistence.

REFERENCES

G. Bittlingmayer (1990), "Efficiency and entry in a simple airline network," International Journal of Industrial Organization, forthcoming.

D. Carlton and J. M. Klamer (1983), "The need for coordination among firms with special reference to network industries," University of Chicago Law Review, 50, 446-465.

G. Faulhaber (1975), "Cross subsidization: pricing in public enterprises," American Economic Review," 65, 966-977.

E. Kalai and E. Zemel (1982), "Generalized network problems yielding totally balanced games," Operations Research, 30:5, Sept-Oct.

D. Granot and G. Huberman (1981), "Minimum cost spanning tree games," Mathematical Programming, 21, 1-18.

D. Granot and M. Hojati (1990), "On cost allocation in communication networks," Networks, 20:2, 209-229.

H. Moulin (1982), Game Theory for the Social Sciences, New York University Press: New York.

L. Shapley (1967), "On balanced sets and cores," Naval Research Logistics Quarterly, 14:4, December, 9-25.

W. Sharkey (1982a), "Suggestions for a game-theoretic approach to public utility pricing and cost allocation," Bell Journal of Economics.

W. Sharkey (1982b), The Theory of Natural Monopoly, Cambridge University Press: Cambridge.

W. Sharkey (1988), "Supportability of network cost functions," Bell Communications Research Economics Discussion Paper #41, July.

ENDNOTES

1 - Sharkey (1982b), p. 184-5.

2 - On the other hand, increased subscription may raise congestion at the nodes and along the links of the network.

3 - Additional monitoring and billing expenses would also arise. If, for instance, services are provided by s suppliers, then $s(s-1)/2$ bilateral accounts must be maintained. For more on transactions costs of networks, see Carlton and Klamer (1983).

4 - In a complete network having n nodes, no fewer than $\sum_1^n n!/(n-m)!$ directed paths run over $n(n-1)$ directed links. As an example, a 5-node network has 20 links and 320 paths.

5 - I conjecture that, when imbedded in more complex networks, they will inherit the whatever instability arises in the simple 3-node network.

6 - See Granot and Hojati (1990).

7 - However, when cost is additively separable across links, and concave on each link, Sharkey (1988) shows that $C(\bullet)$ is concave in undirected flow requirements $\bar{x} = (\bar{x}_{ij})$.

8 - Preferences has the feature of "broadcasting" when $B^n(x_{ij}^n, x_{ik}^n)$. "Shopping behavior" could be described by preferences like $B^n(x_{ij}^n, y_{ij}^n)$.

9 - Because $F_{23} < F_{13}$, if architecture {2,3} is vulnerable to defection by a two-user coalition, then so too is the more costly architecture {1,3}.

10 - Digital, optical communications technologies undergo negligible degradation even if transmissions take roundabout paths. This is not true of analog, electrical systems especially if they involve a satellite link.

FIGURES

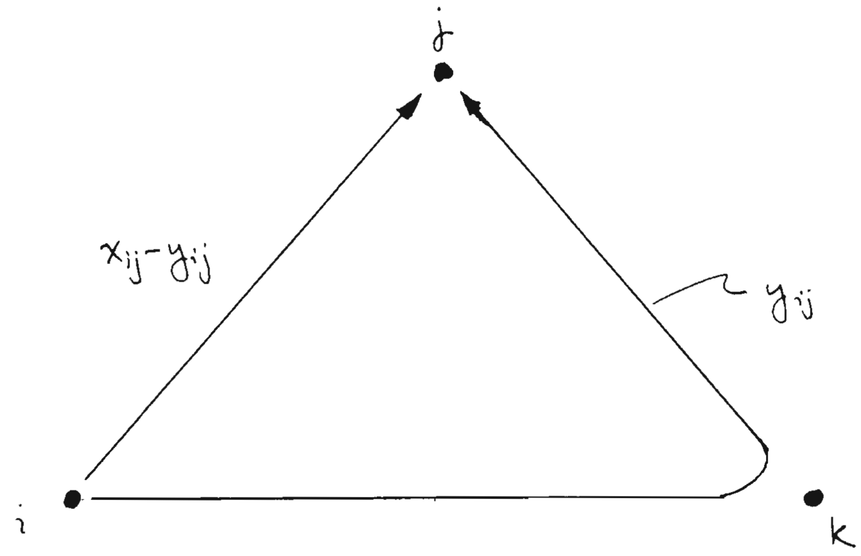

Figure 1: Network Schematic

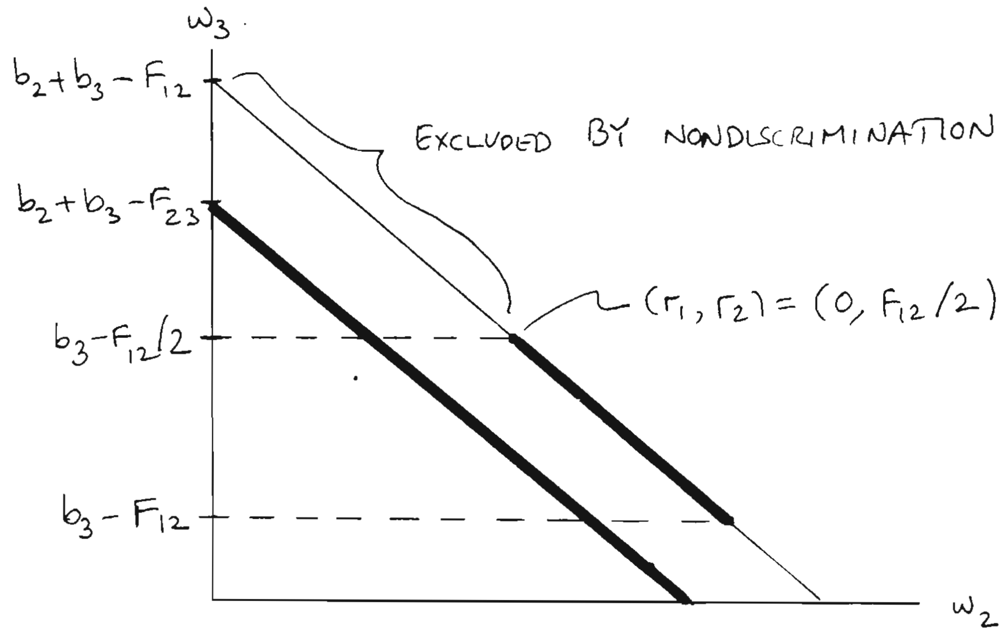

Figure 2: Surplus Possibilities