

CUSTOMER OR COMPLEMENTOR?
Intercarrier Compensation with Two-Sided Benefits

25 June 2009

Benjamin E. Hermalin*
and
Michael L. Katz

We are grateful to Bruno Jullien, Patrick Rey, two referees, a co-editor, and participants at the 2004 IOfest and the 2005 European Summer Symposium in Economic Theory for helpful comments and suggestions.

* Haas School of Business, University of California, Berkeley.

** Stern School of Business, NYU, and Haas School of Business, University of California, Berkeley.

Abstract

Both senders and receivers of telecommunications messages derive benefits from those messages. Hence, consumers' decisions to send and receive messages generate external effects for each other. When a sender and a receiver are both customers of a single carrier, that carrier has incentives to internalize these effects. We explore the use of an access charge (whereby the sender's carrier compensates the receiver's carrier for completing the message exchange) to induce carriers to internalize these external effects when a sender and receiver are customers of different carriers. We find that an access charge can induce constrained-efficient sending and receiving prices for exchanging messages across networks *taking the sum of these prices as given*, but the access charge cannot induce the correct sum. We also show that the internalizing role for the access charge can imply a non-zero access charge is efficient even in highly symmetrical situations, and that the efficient access charge may be positive or negative.

Key Words

Access charge, intercarrier compensation, interconnection, two-sided markets.

I. INTRODUCTION

The benefits from subscribing to a communications network derive from being able to exchange messages with other parties. The interconnection of distinct networks allows users to communicate with a large community of users without the need for carriers to duplicate one and others' networks. Interconnection can thus significantly affect efficiency and market structure. Consequently, rules regarding the terms of interconnection constitute one of the most important areas of public policy concerning telecommunications markets.¹ A critical subset of these rules applies to pricing. When the sender and receiver of a message are customers of two different carriers, questions arise as to whether one carrier should pay the other for its role in facilitating the message exchange. Public policies regarding such intercarrier compensation have been very contentious in both the United States and Europe. As we will discuss, one source of that contentiousness has been the lack of sound theoretical underpinnings for the choice of intercarrier compensation rates.

Consumption of communications services (*e.g.*, talking on the phone, exchanging e-mails, sharing files, or holding a video conference) generally involves a sender and receiver, both of whom take actions, bear costs, and derive benefits.² Until very recently, almost all theoretical work on interconnection pricing ignored the benefits enjoyed by the receiving party.³ This treatment typically was justified by assuming either that the receiving party enjoys no benefits from a message exchange or that the effects between two parties are internalized. The

¹ For a discussion of some of the policy issues, see Federal Communications Commission (2005).

² The fact that multiple parties consume a single message gives rise to external effects. See Hermalin and Katz (2001b), Laffont and Tirole (2000), and Taylor (1994, Chapter 9) for surveys of telecommunications externalities.

³ Leading analyses without receiver benefits include Armstrong (1998) and Laffont et al. (1998a and b).

first assumption clearly is unrealistic. Were it correct, we would never answer the telephone or read our e-mail. The second assumption is applicable only to situations in which either the two parties are altruistic with respect to one another or have a repeated relationship.⁴

Recognition that both sender and receiver enjoy benefits has important implications for efficient pricing to end users.⁵ These, in turn, have further implications for the efficient pricing of interconnection. In the absence of receiver benefits, the sender can be viewed as the “cost causer.” This view suggests that the receiver’s network should recover its message costs from the sender, either directly by billing the sender, or indirectly by billing the sender’s carrier (*i.e.*, by levying an access charge). Traditional pricing of long-distance telephone calls reflects this perspective—the party initiating a call typically is the only one charged for it. Similarly, many countries other than the United States have calling-party-pays pricing for mobile telephony. That is, when a party calls a mobile phone user, the calling party’s carrier is charged a fee by the receiver’s carrier, and this fee is passed on to the calling party, while the called party does not pay. The view that the originating carrier is purchasing services from the terminating carrier also forms the basis for telecommunications policy in the United States. According to the Federal Communications Commission, “under the existing regimes, the calling party’s carrier ... compensates the called party’s carrier for terminating the call.”^{6,7}

⁴ Loosely speaking, Willig (1979, pages 124–25) shows that, if sending a message triggers a set number of incoming messages, then receiver benefits will be internalized in the demand for sending messages. Hermalin and Katz (2001b, §3.5) develop a simple game-theoretic model in which users can (partially) internalize external benefits by engaging in tit-for-tat message initiation.

⁵ For a discussion of the earlier literature on retail pricing in the presence of two-sided benefits, see Hermalin and Katz (2004) and references therein.

⁶ Federal Communications Commission (2005).

⁷ A recent draft decision of the European Commission (undated, ¶14) acknowledges that both parties to a call enjoy benefits from the message exchange but then largely ignores the implications of that fact.

In the presence of receiver benefits, the standard rationale for access charges—that the calling party is the “cost causer”—makes little sense. One could just as well assert that the receiver causes the costs by accepting the message.⁸ Instead of viewing the originating carrier as a customer purchasing terminating access services, one could just as well think of the terminating carrier as purchasing origination services. It is better, however, to think of each carrier as providing a service that is complementary to the other. An important implication of this reframing of the transaction in this way is that there is no *a priori* reason to believe that it is efficient for the calling party’s carrier to compensate the called party’s carrier for terminating the call.

In the present paper, we examine the structure of an efficient inter-carrier compensation regime recognizing that both senders and receivers benefit from electronic message exchange. Our analysis proceeds as follows. We first characterize the socially optimal end-user prices in the presence of two-sided benefits, and we extend the literature (specifically, Hermalin and Katz, 2004, and Laffont et al., 2003) by allowing for hookup fees (*i.e.*, fixed charges for network subscribership). Because prices can play a role in internalizing external effects across the two parties to a message exchange, it is inefficient to have one party bear the full marginal costs of exchanging a message, as under the calling-party-pays model. Indeed, the efficient send and receive prices generally sum to less than marginal cost.

⁸ In the document quoted in the text, the Federal Communications Commission went on to say that “Developments in the ability of consumers to manage their own telecommunications services undermine the premise that the calling party is the sole cost causer and should be responsible for all the costs of a call.” Although the Commission is correct that is invalid to presume that the calling party is the sole cost causer, this fact has nothing to do with recent developments.

We then explore the relationship between access charges and equilibrium end-user prices. A network operator has incentives to implement efficient, below-cost send and receive prices when it is able to capture some or all of the increased consumption benefits in the form of higher network subscription fees. If a sender and receiver subscribe to the same network (a situation known as on-net message exchange), then the network operator is able to capture at least some of the incremental benefits associated with internalizing what would otherwise be consumption externalities across the two end users. In our model, in which the capture is complete, the network operator sets the send and receive prices for on-net calling at their efficient levels. However, if the sender and receiver are on different networks (a situation known as off-net message exchange), then the benefits from subsidizing one side of a message exchange accrue, in part, to a subscriber on another network. When each network operator can collect revenues solely from its own subscribers (as is typically the case) and there is no other coordination or internalization mechanism across networks, there is no way for a network operator to benefit from internalizing the externalities associated with off-net message exchange.⁹ Hence, each network sets its off-net prices to maximize the sum of the net benefits enjoyed by the network operator and subscribers to its network. The resulting prices are set equal to marginal cost, where marginal cost includes any access charge payments or receipts. In summary, network operators in our model set the prices for on-network message exchange at efficient levels, but set the prices for off-net message exchange inefficiently high. Consequently, what Laffont, Rey, and Tirole (1998b) label tariff-mediated network effects arise.

⁹ In a more general model, it is conceivable that a carrier's subsidizing subscribers to a rival carrier would so increase the number of end users connected to the overall network that the value of the subsidizing carrier's own service would rise by enough to make the subsidy profitable. Intuitively, such effects are unlikely to be significant when there are several different interconnected carriers.

We show below that the use of access charges cannot induce off-net efficiency except in a constrained sense. An increase in the access charge lowers the equilibrium price for receiving an off-net message, but it raises the price of sending one. The sum of these prices remains equal to the sum of the marginal costs, which is inefficiently high in the presence of two-sided benefits. One might suspect that access charges would be irrelevant in this setting. However, as we show, access charges still can have efficiency effects by affecting the allocation of marginal cost between the end-user prices for sending and receiving messages across networks. We show that, even in what appear to be very symmetrical situations, a non-zero access charge can be efficient. In stark contrast to analyses that ignore receiver benefits, we show that the optimal access charge can be positive or negative.

The remainder of the paper is organized as follows. Section II describes the model, and Section III characterizes end-user prices. Section IV characterizes the equilibrium outcome taking the access charge as given. The results of this analysis are then used in Section V to derive the optimal access charge. Section VI describes the relationship between the present paper and previous work on access charges in the presence of two-sided benefits. The paper closes with a short summary.

II. A MODEL

We model communication between different end users as a message exchange. A message might be a telephone call, a paging message, an SMS message, a data file, or an e-mail. One party initiates the communication (*e.g.*, places a phone call) and the other party accepts it (*e.g.*, answers the phone). We refer to the initiator as the *sender* and the acceptor as the *receiver*.

When message exchange generates benefits and costs for both parties, there are important differences between situations in which either party can initiate a message exchange (“two-way

calling”) and those in which only one party can do so (“one-way calling”). Our analysis, like the rest of the literature on intercarrier compensation with two-sided benefits, is restricted to one-way calling models.

One-way calling has several interpretations. One is that message origination is literally one-sided. Many telecommunications technologies, such as paging and pay phones are inherently one-way technologies. Other technologies are two way, but in many instances only one of the two parties knows there is value in communicating. For instance, *A* could wish to announce some news to *B*. Alternatively, *A* could be a consumer calling a pizza parlor, *B*, to order a pie. Or, *A* could be an end user establishing a dial-up connection with her Internet service provider, *B*. In such situations, it is reasonable to view only one of the two parties as the potential message initiator. Other situations, in which the parties both know there’s a value to communicating and it is technically feasible for either party to initiate a message exchange, are two-way calling situations.¹⁰

There are two networks, 1 and 2, which supply communication services that are homogenous except for any network effects. A network incurs cost c to originate or terminate a

¹⁰ An alternative interpretation of the distinction between one-way and two-way calling models is the following. With a two-way technology, if a message costs more to send than receive, then a party may strategically delay sending a message in anticipation that the other party will initiate the exchange instead. For cheaply priced messages, this type of strategic behavior may be implausible, and the situation can be approximated by one-way calling model in which a party sends message whenever her expected value of message exchanges exceeds the price she must pay. There is, however, a vexing modeling issue. Suppose that send prices are greater than receive prices. Then a carrier might increase its subscribers’ welfare by raising its off-net send price so that a subscriber would be the receiver a greater percentage of the time. Would a non-strategic end user recognize such effects? And if not, how would he or she evaluate alternative networks when making subscription decisions? One possibility would be for end users to have long- and short-run selves, but this sort of modeling goes well beyond the scope of the present paper.

message, and the incremental cost of a message exchange is thus $2c$ regardless of the identities of the originating and terminating networks.¹¹

Network i offers a multipart tariff with a fixed hookup fee, h_i , and a set of traffic-sensitive charges. Let p_{ij} denote the retail price charged to a user on network i who sends a message to a user on network j . We refer to p_{ij} as the on-net send price when $i = j$ and the off-net send price when $i \neq j$. Define the receive prices, r_{ij} , as the price charged to a user on network i who receives a message from a user on network j . Henceforth, it is to be understood that $i \neq j$, so that, for example, p_{ii} and p_{ij} denote the on-net and off-net send prices, respectively, charged by network i .

There are $\bar{N} < \infty$ end users who are potential subscribers to the two carriers. A given end user subscribes to the services of at most one carrier. Any given user, k , has probability λ of wanting to initiate a message exchange with any other end user, ℓ . We assume that λ is independent of k and ℓ . Given this assumption, there is no further loss of generality from setting $\lambda = 1$, which simplifies the notation below. Let ω denote the value of a message sent or received, and let $\psi(\omega)$ denote the density for the value of message exchange conditional on the sender's wanting to initiate a message exchange.¹² We assume that the values of message exchange are distributed independently across potential exchange partners. Without loss of

¹¹ As will become evident below, the model readily generalizes to settings in which: (a) the marginal costs of origination and termination differ, and (b) there is an additional marginal cost of intermediate transport.

¹² Alternatively, one could define ω as the *expected* value of message exchange because the realized (*ex post*) value of a message could be unknown *ex ante*; that is, sender and receiver could each learn, after message exchange, that the actual value of a message was different than they expected when they chose to send and receive the message. See Hermalin and Katz (2004) for a discussion. In the context of the present paper, the distinction between *ex ante* and *ex post* values is unimportant, and one can adopt either interpretation.

generality, we normalize message values so that the support of $\psi(\cdot)$ has an infimum of 0. To avoid trivial cases, we assume throughout that the supremum of the support of $\psi(\cdot)$ exceeds c and that the mean value of ω is less than $2c$.

Define $D(q) \equiv \int_q^\infty \psi(\omega) d\omega$. $D(q)$ is the probability that a user's expected value of a message exchange exceeds q . $D(p)$ is the probability that any given end user will wish to call any other particular end user when the send price is p , and $D(r)$ is the probability that any given end user will wish to receive a call from any other given end user when the receive price is r . Next, define $S(q) \equiv \int_q^\infty (\omega - q)\psi(\omega) d\omega$. $S(q)$ is the expected surplus of a message (sent or received) for which the user pays q . Lastly, define $M(q)$ as the mean of ω conditional on $\omega \geq q$ and observe that:

$$M(q) = \frac{S(q)}{D(q)} + q . \quad (1)$$

Each user is motivated only by his or her private net benefits, and income effects are assumed to be zero. An end user who subscribes to network i enjoys expected surplus

$$\tau + (n_i D(r_{ii}) S(p_{ii}) + n_j D(r_{ji}) S(p_{ij})) + (n_i D(p_{ii}) S(r_{ii}) + n_j D(p_{ji}) S(r_{ij})) - h_i , \quad (2)$$

where τ is the non-network component of the consumer's willingness to pay and n_k is the number of subscribers to network k . The terms in large parentheses are the expected surplus from initiating and receiving calls, respectively. τ is the standalone value to the carrier's service, which may be positive or negative. A positive standalone value could represent the benefits of having access to emergency services, such as when one dials 911 in the United States or 999 in the United Kingdom. Negative values can correspond to the opportunity cost of time spent subscribing to the network or the costs of any equipment that an end user has to purchase

to make use of the network. One can also interpret τ and the linear network effects as an affine approximation to a concave network benefit function.

Define $T(n)$ as the value of τ for the end user with the n^{th} highest value. That is, there are n end users for whom $\tau \geq T(n)$. For convenience, we treat n as a continuous variable. We assume that: (a) $T'(n) < 0$ and $T''(n) < 0$ for all n , and (b) $T(n) \downarrow -\infty$ as $n \uparrow \bar{N}$.

We examine the pure strategy equilibria of the following game:

- The access charge, a , that carrier i must pay to carrier j for each message that originates on i and terminates on j is specified. We assume that the access charge is the same for both carriers (*i.e.*, it is a symmetric reciprocal compensation scheme). We analyze the choice of access charge made by a total-surplus-maximizing regulator and by the carriers themselves. We observe that, *a priori*, the socially optimal access charge could be negative or positive.
- Carriers simultaneously choose their numbers of subscribers, n_1 and n_2 .
- Given the carriers' quantity choices, prices adjust to clear the market. The presence of multi-dimensional pricing strategies raises a complication that does not arise in the standard Cournot model.¹³ However, as we show below, there is a strong sense in which each carrier's pricing decision reduces to a single dimension.
- Having made his or her network subscription decision, each end user chooses whether to participate in a message exchange. We impose a perfection requirement that each

¹³ For an early use of the Cournot assumption in a model of network effects (with uni-dimensional pricing), see Katz and Shapiro (1985).

consumer maximizes his or her net benefits in the final stage taking prices and network sizes as given.

III. END-USER PRICES

Although our focus is on inter-carrier pricing, the welfare consequences of interconnection charges are mediated through the effects of access pricing on the resulting retail prices charged to end users.¹⁴ It is thus useful to begin by characterizing certain end-user prices.

A. Efficient and Constrained-Efficient Prices

Consider, first, efficient end-user prices. As will become evident below, because we analyze a Cournot model, we can limit our examination of efficient end-user prices to situations in which network sizes are given. Conditional on network subscribership levels, the expected surplus realized by an end user who subscribes to network i is given by (2) above.

Conditional on network subscribership, the efficient prices are those that maximize calling benefits net of costs, or

$$\sum_{i=1}^2 n_i \left\{ n_i [D(p_{ii})S(r_{ii}) + D(r_{ii})S(p_{ii}) + (p_{ii} + r_{ii} - 2c)D(p_{ii})D(r_{ii})] \right. \\ \left. + n_j [D(p_{ij})S(r_{ji}) + D(r_{ji})S(p_{ij}) + (p_{ij} + r_{ji} - 2c)D(p_{ij})D(r_{ji})] \right\} .$$

We assume that the gross benefits of exchange are large enough relative to costs that shutting down both networks is not efficient. Given our assumption that the benefits of exchange are bounded below by 0, setting either the send or receive price below 0 offers no efficiency gains; hence, we may take these prices as being bounded below by 0.

¹⁴ In other settings, intercarrier compensation could also affect network investment decisions. For analyses, see Hermalin and Katz (2006) and Valletti and Cambini (2005).

Observe that maximizing total surplus can be divided into four independent maximization problems of the form

$$\max S(p)D(r) + S(r)D(p) + (p + r - 2c)D(p)D(r).$$

The derivative with respect to the send price is

$$D'(p)\{S(r) + (p + r - 2c)D(r)\}.$$

Because user surplus and demand are positive, the change in welfare due to a marginal increase in the send price is negative if $p + r \geq 2c$. Applying a similar logic to the choice of the receive price, it follows that any efficient price pair, say (p^*, r^*) , must satisfy $p^* + r^* < 2c$.¹⁵ Intuitively, each end user's decision to participate in a message exchange generates positive expected surplus for the other party to the exchange and below-cost pricing serves to internalize this effect. Because efficient send and receive prices do not cover costs, optimal pricing subject to a carrier breakeven constraint entails a strictly positive hookup fee. End-user homogeneity with respect to the distribution of ω implies that subscribers' individual rationality constraints will be satisfied by efficient pricing.

We can use equation (1) to rewrite the derivative with respect to the send price as

$$D'(p)D(r)\{M(r) + p - 2c\}. \quad (3)$$

Given an r for which $D(r) > 0$, an interior solution for p exists if and only if $M(r) - 2c \leq 0$ and a corner solution of $p = 0$ exists if and only if $M(r) > 2c$. Intuitively, p is the value of a marginal message to the sender who faces price p , and $M(r)$ is the expected gross benefit enjoyed by a receiver who accepts the message conditional on its value being greater than r . Hence, the first-order condition indicates that, if it is an interior solution, p must be set so that the expected gross

¹⁵ This result extends Hermalin and Katz's (2004) Proposition 1 to the case of multiple networks.

consumption benefits of the marginal message, $p + M(r)$, are equal to the marginal cost, $2c$.¹⁶

Similar considerations apply to the choice of the receive price.¹⁷

In the Appendix, we prove a further characterization of the efficient prices:

Proposition 1: *Suppose hook-up fees are feasible.*

- (i) *If the hazard rate associated with $\psi(\cdot)$ is everywhere increasing, then there is a unique pair of send and receive prices that maximize total surplus and they satisfy $0 < p^* = r^* < c$.*¹⁸
- (ii) *If the hazard rate is constant, then any prices such that $0 \leq p \leq 2c - M(0)$ and $r = 2c - M(0) - p$ are socially optimal.*
- (iii) *If the hazard rate is everywhere decreasing, then there are two socially optimal price pairs: one in which the send price equals 0 and one in which the receive price equals 0. In each case, the complementary price equals $2c - M(0) > 0$.*

For subsequent analysis, it is also useful to characterize the efficient prices when there is a per-message breakeven constraint, $p + r = 2c$. In an earlier article, we showed:¹⁹

Hermalin and Katz (2004), Proposition 2: *Suppose transaction prices are subject to a breakeven constraint. Consider the hazard rate associated with density $\psi(\cdot)$.*

¹⁶ In a paper written independently of ours, Bolt and Tieman (2005) derive a similar necessary condition for the socially optimal prices of a monopolist that is not subject to a profitability constraint and that cannot charge a hookup fee.

¹⁷ The arguments in the text above and in the proof of Proposition 1 apply for any n_1 and n_2 ; given the efficient send and receive prices, socially optimal pricing entails adjusting the hook-up fee to ensure that end users want to subscribe to the networks.

¹⁸ In an insightful paper, Laffont et al. (2003, Proposition 2) derive a similar result but make an error. They do not check the second-order conditions. Doing so reveals that their first-order condition is neither necessary nor sufficient. Consequently, Laffont et al. do not identify possibilities (ii) and (iii).

¹⁹ We have adapted the statement of our original proposition to fit the notation of the current analysis.

- (i) *If the hazard rate is everywhere increasing, then prices that divide the cost of a message equally between the sender and receiver (i.e., $p = c = r$) are the unique socially optimal prices.*
- (ii) *If the hazard rate is constant, then any prices such that $0 \leq p \leq 2c$ and $r = 2c - p$ are socially optimal.*
- (iii) *If the hazard rate is everywhere decreasing, then there are two socially optimal price pairs: one in which the send price equals 0 and one in which the receive price equals 0. In each case, the complementary price is set at $2c$.*

These results illustrate the well-known fact that the first- and second-best prices depend only on the sum of the origination and termination costs. Hence, our assumption that the two costs are equal is inconsequential.

Lastly, it is useful to define p^c and r^c as the prices that maximize the sum of profits and consumer surplus that carrier i and its subscribers derive from off-net calling. Given the separability of the sending and receiving programs and the fact that consumer surplus and profits are linear in the number of messages exchanged, the prices p^c and r^c are the solutions to:

$$\max_p S(p) + (p - c - a)D(p)$$

and

$$\max_r S(r) + (r - c + a)D(r)$$

Assuming that $0 \leq c - |a| \leq c + |a| \leq \sup \text{supp } \psi(\cdot)$, the solution is readily shown to be

$p^c = c + a$ and $r^c = c - a$. This finding is intuitive: a carrier engages in marginal cost pricing, which maximizes the net benefits accruing to that carrier and its subscribers while ignoring any effects on the rival network and its subscribers.

In summary, the efficient send and receive prices take into account the effects on both the sender and the receiver and on any carrier involved. The off-net prices just derived ignore the

effects of prices on the welfare of consumers who subscribe to another carrier and on the profits of that other carrier.

B. Market-Clearing Prices

Given any $n_1 > 0$ and $n_2 > 0$, consumers are willing to join the two networks in the required numbers if and only if

$$n_i(D(p_{ii})S(r_{ii}) + D(r_{ii})S(p_{ii})) + n_j(D(p_{ji})S(r_{ij}) + D(r_{ji})S(p_{ij})) - h_i + T(n_1 + n_2) = 0 \quad (4)$$

for $i = 1, 2$. Equation (4) requires that the two networks offer the same level of consumer surplus to any subscriber and that the marginal subscriber be indifferent between subscribing to a network and forgoing the service entirely. For $n_i = 0$, the left-hand side of (4) must be less than or equal to 0.

We have a situation with two equations and ten unknowns, and there are many solutions to (4) for any given pair of subscriber levels. However, as we now show, we can narrow down the set of solutions for each firm to a single-parameter family of prices. Specifically, we show that the profit-maximizing send and receive prices are independent of the carriers' choices of subscriber levels. Because a carrier can use the hookup fee to extract surplus from its subscribers, a carrier can fully internalize on-net network effects. However, because a carrier cannot collect revenues from subscribers to its rival's network, a profit-maximizing carrier does not internalize off-net effects (except through the access charge, as discussed below). Consequently, a carrier's profits are always maximized by setting its on-net prices at p^* and r^* and its off-net prices equal to p^c and r^c :

Proposition 2: *Given any n_1, n_2 , and prices charged by the other network, network i earns greater profits charging the market-clearing price vector with on-net prices p^* and r^* and off-net prices p^c and r^c than it earns charging any other market-clearing prices.*²⁰

In the light of Proposition 2, we take the market-clearing prices to be those stated in the hypothesis of the result. Thus, as in the standard Cournot model, one can think of there being a well-defined pair of market-clearing prices (*i.e.*, the hook-up fees) given any pair of supplier output levels. Given any $n_1 > 0$ and $n_2 > 0$, consumers will be willing to join the two networks in the required numbers if and only if

$$h_i = n_i u^* + n_j w^c + T(n_1 + n_2)$$

for $i = 1, 2$, where $u^* \equiv D(p^*)S(r^*) + D(r^*)S(p^*)$ and $w^c \equiv D(p^c)S(r^c) + D(r^c)S(p^c)$. Notice that, because off-net prices are set equal to the constant marginal cost, the level of consumer surplus derived from off-net calls is equal to the total surplus derived from those calls.

IV. EQUILIBRIUM

Given n_j , network i chooses n_i^r to maximize

$$n_i^r \left\{ h_i + n_i^r (p^* + r^* - 2c) D(p^*) D(r^*) \right\} = n_i^r \left\{ n_i^r u^* + n_j u^c + T(n_i^r + n_j) + n_i^r \pi^* \right\},$$

where $\pi^* \equiv (p^* + r^* - 2c) D(p^*) D(r^*) < 0$.²¹ π^* is the amount that an on-net call is subsidized times the probability that one on-net end user calls another and has that call answered. Let w^* denote the expected total surplus generated when one end user calls another end user subscribing

²⁰ The prices in the proposition yield network i strictly greater profits than any other prices whenever there is off-net traffic (*i.e.*, as long as both networks have subscribers and network j does not set its off-net prices so high as to cut-off all off-net message exchange).

to the same network; that is, $w^* \equiv u^* + \pi^*$. The first-order necessary condition for carrier i 's maximization problem is

$$2n_i^r w^* + n_j w^c + T(n_i^r + n_j) + n_i^r T'(n_i^r + n_j) \leq 0, \quad (5)$$

with strict equality if $n_i^r > 0$.

Depending on parameter values, several types of equilibrium may exist.²² We consider each, in turn.

Proposition 3: *There exists a symmetric duopoly equilibrium with positive sales if and only if there exists a positive number of subscribers, N^d , such that*

$$0 \leq \frac{1}{2} N^d \{w^* + w^c\} + T(N^d) = -\frac{1}{2} N^d \{w^* + T'(N^d)\}. \quad (6)$$

There exists at most one symmetric equilibrium with positive sales.

We can relate the existence of symmetric equilibria with positive sales to simple conditions on the distribution of standalone benefits:

Corollary: *If $T(0) > 0$ and $w^* + T'(0) \leq 0$, then there exists a symmetric equilibrium with positive sales. If $T(0) \leq 0$ and $\frac{1}{2}(w^* + w^c) + T'(0) \leq 0$, then there does not exist a symmetric equilibrium with positive sales.*

It is readily shown by example that, when $T(0) \leq 0$ and $\frac{1}{2}(w^* + w^c) + T'(0) > 0$, there exists a symmetric equilibrium with positive sales for some functional forms but not others.

Next, consider the possibility of asymmetric equilibria. Suppose that only one carrier attracts subscribers. If that network has N subscribers, then its profits are equal

²¹ There can be multiple values of p^* and r^* in some cases. What matters for our analysis, however, is w^* , which is unique and greater than w^c .

to $N\{Nw^*+T(N)\}$. If no N exists for which $Nw^*+T(N) > 0$, then the monopolist can do no better than to shut down. Suppose such an N does exist. By assumption, there exists some $N_0 < \bar{N}$ such that $Nw^*+T(N) < 0$ for all $N > N_0$. Hence, one can view the firm as choosing the most profitable value of N from the compact interval, $[0, N_0]$, and an optimum exists by application of Weierstrass's Theorem. The first-order necessary conditions ((5) above) for the active and inactive carriers are $2Nw^*+T(N) + NT'(N) = 0$ and $Nw^c + T(N) \leq 0$, respectively.

We have established:

Proposition 4: *There exists a monopoly equilibrium with positive output if and only if there*

exists a number of subscribers, N^m , such that $w^ > -\frac{T(N^m)}{N^m} \geq w^c$ and*

$$-\frac{T(N^m)}{N^m} - T'(N^m) = 2w^* .$$

Lastly, for some parameter values, asymmetric interior equilibria can exist. By (5), such an equilibrium exists if and only if :

$$n_1\{2w^*+T'(n_1+n_2)\} + n_2w^c = -T(n_1+n_2)$$

and

$$n_2\{2w^*+T'(n_1+n_2)\} + n_1w^c = -T(n_1+n_2)$$

for some $n_1 \neq n_2$. These two conditions can be satisfied for $n_1 \neq n_2$ only if

$$2w^* - w^c = -T'(n_1+n_2) . \tag{7}$$

By the concavity of $T(\cdot)$, (7) is satisfied by at most one value of $n_1 + n_2$. For the case in which such a solution exists, label the sum of subscriber levels \hat{N} . Summing the two first-order conditions implies that we must have

$$\hat{N}\{2w^* + w^c\} = -2T(\hat{N}) - \hat{N}T'(\hat{N}). \quad (8)$$

Together, (7) and (8) imply

$$\hat{N}\{2w^* + w^c\} = -2T(\hat{N}) + \hat{N}\{2w^* - w^c\}$$

or

$$\frac{-T(\hat{N})}{\hat{N}} = w^c. \quad (9)$$

Summarizing this analysis,

Proposition 5: *If there exists an industry subscription level, \hat{N} , such that $-T'(\hat{N}) = 2w^* - w^c$*

and $\frac{-T(\hat{N})}{\hat{N}} = w^c$, then there exists a continuum of equilibria subscription levels of the form

$(n_1, \hat{N} - n_1)$ for $n_1 \in [0, \hat{N}]$. If these conditions are not satisfied, then there are no interior asymmetric equilibria.²³

V. OPTIMAL ACCESS CHARGES

To understand the effects of the access charge on welfare, we need to examine how the access charge affects the equilibrium outcome through its effects on w^c (recall that w^* and $T(\cdot)$ are independent of a). In principle, there are potentially two roles for an access charge to

²³ If $T(0) \geq 0$, then $-T(\hat{n})/\hat{n} < -T'(\hat{n})$. Hence, in this case, asymmetric equilibria can exist only if $2w^* > 3w^c$.

play. One is to induce pricing below cost in order to internalize the effects of calling externalities. Unfortunately, Proposition 2 implies that an access charge cannot serve this purpose because the access charge has no effect on the sum of the equilibrium off-net prices: $p_{ij} + r_{ji} = 2c$ whenever there is off-net traffic. In other words, the access charge cannot be used to internalize the effects of a carrier's prices on the welfare of consumers who subscribe to another carrier and on the profits of that other carrier. The carriers can be induced to set efficient off-net prices only if there some mechanism for cross-carrier internalization such as repeat play or strategic delegation in the manner of Katz (2006), which allow each carrier to make its off-net pricing policies contingent on the other carrier's off-net pricing policies.

Let \bar{w}^c be the largest value of w^c that can be induced through the choice of a . Because the off-net send and receive prices sum to $2c$, which is inefficiently large, $\bar{w}^c < w^*$ except in cases where all of the probability mass is on sending and receiving values greater than c . Of course, the fact that access charges cannot promote fully efficient pricing does not imply that they are irrelevant. Access charges can be used to induce a balance of off-net send and receive prices that is efficient given the constraint that these prices sum to $2c$.

As shown in the previous section, there are potentially three types of equilibrium to consider. The following result demonstrates that a regulator would never want to induce an interior asymmetric equilibrium:

Proposition 6: *A socially optimal access charge induces either the monopoly outcome, or the symmetric duopoly outcome.*

As long as the market outcome has only a single active producer, the value of the access charge is irrelevant. However, when the industry equilibrium involves two active networks,

welfare depends on the value of w^c , which, in turn, depends on the value of a . The next result tells us that, conditional on inducing a symmetric duopoly equilibrium, it is socially optimal to choose the access charge that induces \bar{w}^c .

Proposition 7: *Suppose that there exists a positive-output, symmetric equilibrium when $w^c = w_L$. Then for any $w_H > w_L$, there exists a symmetric equilibrium when $w^c = w_H$ in which the equilibrium network sizes, profits, consumer surplus, and, hence, total surplus are greater than when $w^c = w_L$.*

Observe that one implication of Proposition 7 is that, when both are active, the carriers also prefer the access charge that maximizes w^c . Combined with Proposition 6, another implication of Proposition 7 is that it is socially optimal either to select the monopoly outcome or to choose a level of the access charge that induces \bar{w}^c and select the duopoly outcome. In what follows, we will assume that there exists at least one value of a such that a monopoly equilibrium exists.²⁴

Proposition 8: *Suppose $T(0) \geq 0$. If \bar{w}^c is sufficiently close to w^* , then the symmetric duopoly equilibrium with the access charge set to induce \bar{w}^c is the socially optimal outcome. If \bar{w}^c is sufficiently low, then the monopoly equilibrium is the socially optimal outcome.*

The public policy implications of our analysis are as follows. Policy makers should first calculate whether the monopoly or duopoly outcome is more desirable. If the duopoly outcome is preferable, then—as long as higher values of w^c do not increase the probability that the

²⁴ As long as the mean values of message exchange are finite, it is possible to induce $w^c = 0$ by choosing an extreme value of a .

market will select the monopoly equilibrium—policy makers should choose a to maximize \bar{w}^c . Conditional on the duopoly equilibrium's obtaining, welfare is maximized by maximizing w^c , and, if the monopoly outcome were to result, the choice of a would have no effect. Moreover, as the analysis shows, for high enough values of w^c , a monopoly equilibrium does not exist; that is, by inducing \bar{w}^c , policy makers may ensure the desired equilibrium will be the one to arise.

A parallel argument does not call for minimizing \bar{w}^c when monopoly is socially preferred. The asymmetry arises because the value of a matters if the market selects the duopoly equilibrium despite the social preference for monopoly. Moreover, although a low value of w^c can preclude the existence of a duopoly equilibrium, it need not. In cases in which the monopoly outcome is the socially preferred outcome, there could be a role for tools other than the access charge (*e.g.*, granting exclusive licenses) to play in promoting the desirable equilibrium selection.²⁵

The following two examples demonstrate that either the monopoly or duopoly outcome can be socially preferred, depending on the parameter values.

Example 1: Suppose $T(0) \geq 0$ and that almost all of the mass for the value of message exchange is on values greater than c . Then there is vanishingly small efficiency benefit of setting sending or receiving prices below cost and \bar{w}^c can be made arbitrarily close to w^* , where the maximal value of w^c can be obtained by setting $a = 0$. It is readily shown that the symmetric duopoly equilibrium when w^c is near w^* is welfare superior to the monopoly

²⁵ That said, the granting of exclusive licenses suffers from a serious shortcoming: if policy makers have underestimated the value of competition, there is no market-based “safety valve” to override their mistaken preference for monopoly.

outcome because duopoly garners additional subscribers and per-subscriber message exchange is almost equally efficient.

Example 2: Suppose $D(q) = v - q$ and, hence, $S(q) = \frac{1}{2}(v - q)^2$, where v is a constant satisfying $4c > v > c$. By Proposition 1 above, $p^* = r^*$. First-order condition (3) above implies that $p^* = \frac{1}{3}(4c - v) = r^*$. By Proposition 2 of Hermalin and Katz (2004), $p^c = r^c$ at the value of a that induces $w^c = \bar{w}^c$. Straightforward calculations reveal that $\bar{w}^c = \frac{27}{32} w^*$. Suppose $T(N) = -N^{7/6}$.²⁶ Then, by (6), $T(N^d)/N^d = -\frac{273}{304} w^*$ when $w^c = \bar{w}^c$. As shown in the proof of Proposition 8 in the Appendix, the sign of $N^d - N^m$ is equal to the sign of $\bar{w}^c + T(N^d)/N^d$. For the parameters of this example, this expression is equal to $\frac{27}{32} w^* - \frac{273}{304} w^* < 0$. Thus, the symmetric duopoly outcome entails both lower subscriber levels and lower average calling benefits per subscriber.

Ceteris paribus, carrier competition increases subscription levels by the standard Cournot logic that each carrier ignores the fact that its expansion leads to a marginal consumer with a lower willingness to pay and, thus, reduces the revenues of the rival carrier. However, unlike a monopoly carrier, duopolists do not fully internalize the external effects that the sender and receiver of a call have on each other. As just shown, either effect can dominate in terms of its influence on equilibrium welfare.²⁷

²⁶ This function does not asymptotically go to $-\infty$ as N approaches some finite bound. However, it is trivial—though cumbersome—to modify the example to satisfy this property.

²⁷ In a model that allowed for product differentiation and heterogeneous consumer preferences, there would be an additional reason to favor the duopoly outcome over the monopoly outcome. For an early model product differentiation and network effects, see Farrell and Saloner (1986).

The potentially poor performance of duopoly relative to monopoly might appear to be the result of off-net discrimination by competing carriers, which results in $w^c < w^*$. However, a public policy prohibition of off-net discrimination would not eliminate the problem. Absent discrimination, the common on- and off-net prices would not be set at efficient levels. A monopolist fully internalizes the calling benefits; duopolists—even non-discriminating duopolists—do not. Thus, call externalities, rather than tariff-mediated network effects, are the root source of the problem.

In those cases in which the symmetric duopoly outcome is socially optimal, we can use Proposition 2 of Hermalin and Katz (2004) restated in Section III.A above to determine the socially optimal access charge. That earlier result establishes what the socially optimal send and receive prices are when subject to the constraint that these prices sum to marginal cost.

Recalling that the equilibrium off-net prices are $p^c = c + a$ and $r^c = c - a$, one can use the values of the constrained-efficient send and receive prices to determine the optimal access charge:

Proposition 9: *If it induces a symmetric duopoly equilibrium, then the socially optimal access charge depends on the hazard rate associated with density $\psi(\cdot)$ as follows:*

- (i) *If the hazard rate is everywhere increasing, then $a = 0$ is the unique socially optimal access charge.*
- (ii) *If the hazard rate is constant, then any access charge that satisfies $-c \leq a \leq c$ is socially optimal.*
- (iii) *If the hazard rate is everywhere decreasing, then there are two socially optimal access charges: $a = -c$ and $a = c$.*

This result illustrates that it is a mistake to think of the originating carrier as purchasing termination services from the terminating carrier, a view that has long been a foundation of U.S. regulatory policy. Under the conditions of Proposition 9, whenever a_0 is optimal, so is $-a_0$. In other words, one could just as well think of the receiving carrier as purchasing origination services. This result also shows that, even in our highly symmetric model, an access charge of zero need not be optimal

We close this section by considering a restricted form of pricing. Suppose that a carrier can engage in off-net discrimination, but the carrier's on-net send and receive prices must be the same as each other and its off-net send and receive prices must be the same as each other (*i.e.*, prices must satisfy $p_{ii} = r_{ii}$ and $p_{ij} = r_{ij}$). This type of pricing has been practiced by many wireless carriers.²⁸ Under this pricing structure, subscription to network i yields expected consumer surplus

$$\tau + 2(n_i D(p_{ii}) S(p_{ii}) + n_j D(p_{ji}) S(p_{ij})) - h_i$$

Network i 's profits per subscriber are

$$2(n_i D(p_{ii})^2 \{p_{ii} - c\} + n_j D(p_{ji}) D(p_{ij}) \{p_{ij} - c\}) - h_i .$$

The access charge does not appear in this expression for profits because the traffic flows are always balanced: $n_i n_j D(p_{ij}) D(p_{ji})$ messages originate on i and terminate on j , and equally many messages originate on j and terminate on i . Consequently,

²⁸ In the U.S., some wireless carriers offer plans with free on-net calling but costly off-net calling. U.S. wireless carriers also frequently set non-uniform price schedules containing a large number of "free" (on- and off-net, incoming and outgoing) minutes. Once these minutes are exhausted, there is often only a single per-minute charge, regardless of whether a call is on- or off-net and the party is a sender or receiver.

Proposition 10: *Access charges are irrelevant when end users have balanced calling demands and a network's prices are independent of the message direction.*²⁹

VI. RELATIONSHIP TO THE LITERATURE

A small but growing number of papers have analyzed intercarrier compensation in settings with two-sided benefits.^{30, 31} As in our analysis, papers in this literature principally focus on the effects of the access charge on retail prices and resulting consumer behavior.³² These papers vary in terms of assumptions about the distribution of benefits across senders and receivers, the types of pricing allowed, and the nature of duopolistic competition.

Berger (2004), Jeon et al. (2004), and—in his baseline model—DeGraba (2003), all assume that the sender's benefits are a given proportion of the total benefits on every call. One can also view the traditional assumption that only the sender enjoys benefits of message exchange (see, *e.g.*, Laffont et al. 1998a and b) as a polar case of this approach. A critical implication of the constant proportional benefits assumption is that the first-best can be attained by setting send and receive prices that sum to the marginal cost of message exchange. This

²⁹ If there were customers with unbalanced traffic patterns (*e.g.*, a user knew *ex ante* that she was more likely to send a message than to receive one), then a network might offer a two-part tariff that was particularly attractive to a beneficial selection of customers. See, for example, Hermalin and Katz (2001a).

³⁰ In addition to the literature summarized below, Armstrong's (2002) insightful survey of the interconnection literature briefly addresses two-sided benefits in settings where all receivers benefit equally from all messages.

³¹ There are some parallels between telecommunications access fees and electronic payment network interchange fees (the service fee transferred between the buyer and seller's banks when an electronic payment card is used). To draw these parallels, one views a payment card issuer as a carrier for card users and an acquiring bank as a carrier for the merchants at which the card is used. With payment networks, a party typically knows that it will either always be a "sender" or always be a "receiver." Moreover, a given bank may have customers that are all "receivers" or all "senders." Thus, although the general issue of apportioning costs between the two end-users is similar, the particulars of interchange fee setting are different from our model.

implication then has strong implications for optimal access charges. For example, under the assumption that only the sender enjoys calling benefits, the objective of access pricing is to induce a send price equal to the full marginal cost of a call ($p = 2c$ in our notation). When senders and receivers equally split the benefits of message exchange, DeGraba (2003, p. 224) finds that an access charge of 0 can induce efficient retail pricing. Similarly, Jeon et al. (2004, p. 108) find conditions under which the access charge can induce the efficient outcome.

The alternative approach taken in the present paper allows a sender and a receiver's relative valuations of a message to vary across messages. This more realistic assumption leads to very different conclusions. Under this assumption, typically no level of the access charge can induce efficient off-net prices. Moreover, setting the access charge equal to termination cost can be optimal even in highly symmetrical settings.

Our paper shows that the efficient access charge depends on the relative likelihood of disparate sender and receiver benefits versus similar values. This likelihood, in turn, depends on how fat the tails of the sender and receiver benefit distributions are. If the tails are not too fat and the likelihood of very disparate values is, therefore, relatively small, then symmetric send and receive prices are optimal, which argues for $a = 0$. If the tails are fat and the likelihood of very disparate values is relatively great, then unequal send and receive prices are optimal, which argues for $a \neq 0$ although it does not clearly imply whether the access charge should be positive or negative. This insight helps explain the similarity and difference between our work and DeGraba's (2003, Section 4) extension in which both the total value of a message exchange and the division of that value between the sender and receiver can vary by message. DeGraba

³² An exception is Atkinson and Barnekov (2000), who did not consider the effects of access changes on retail prices and conclude that each carrier should recover from its own subscribers all costs not

focuses on the case in which the sender and receiver's benefits are positively correlated and, hence, necessarily considers situations in which the benefits are not disparate, which supports a finding that $a = 0$ is optimal. In the one case with negatively correlated sending and receiving benefits, (DeGraba 2003, Proposition 6 and Figure 1), benefits are necessarily disparate and, accordingly, $a = 0$ is not optimal. Our results do not rely on assumptions about the correlation of benefits—indeed, we assume independent distributions—which helps to show it is the relative likelihood of disparate values rather than correlation per se that is behind the results on the optimal access termination charge.³³

Hermalin and Katz (2001a) and Laffont et al. (2003) also assume that a sender and receiver's relative valuations of a message can vary across messages.³⁴ However, neither paper permits the carriers to have different prices for on-net and off-net message exchange. Consequently, there are no tariff-mediate network effects, and end users' subscription decisions are unaffected by the relative network sizes. Indeed, because both those papers essentially take the subscriber base as given, even the absolute size of the combined networks is irrelevant to end users' subscription decisions.³⁵ In contrast, our focus has been on the consequences that arise when networks can discriminate between on- and off-net messages, including the effect this has on subscription behavior and network size. This difference in approach allows us to identify

incremental to interconnection itself.

³³ Our finding that asymmetric pricing can be optimal when the sender and receiver have identical demand functions also shows that the common approach of recommending that prices be set based on relative price elasticities of demand can be ill-founded.

³⁴ Laffont et al.'s (2003) analysis builds on their characterization of socially optimal retail prices. However, as noted in footnote 18 above, that characterization contains an error, which affects their analysis of socially optimal access charges.

some heretofore unrecognized consequences of the access charge: because it affects off-net pricing only, the access charge affects the strength of tariff-mediated network effects.

Our model also differs from the literature in that we consider the use of hookup fees in addition to on- and off-net discrimination. In a model that does not allow hookup fees, Berger (2004) finds that carriers will use the access charge to facilitate collusion so that the privately optimal access charge will exceed the social optimum. In our model, carriers extract rents through the hookup fee and do not seek to induce inefficiently high prices for off-net message exchange. Consequently, we find that the profit- and total-surplus-maximizing access charges can coincide. Hermalin and Katz (2001a) also do not allow for hookup fees, and Laffont et al. (2003) consider the use of hookup fees but do not allow for differential on- and off-net pricing. In neither of these models are hookup fees used to subsidize efficient, below-cost on-net pricing.

Lastly, consider the modeling of oligopolistic interaction. Berger (2004), Jeon et al. (2004), and Laffont et al. (1998b) each examines a Hotelling duopoly.³⁶ In these models, unlike ours, there is an incentive to degrade the rival carrier's network. Hence, Jeon et al. (p. 108) find that network-based discrimination creates strong incentives for networks to eliminate off-net message exchange even between equal networks. Laffont et al. (1998b) find that there exists a stable duopoly equilibrium only if the access charge is near the termination cost and the carriers are highly differentiated (*e.g.*, the Hotelling transportation costs are large relative to gross

³⁵ Hermalin and Katz (2001a) consider a model with a single sender and a single receiver. In Laffont et al. (2003), a given end user is either always a sender or always a receiver, and senders (websites) pay no hookup or subscription fee and have no overhead costs, so their joining the network is a given. Because all senders connect, receivers do not take network size into account in making their subscription decisions.

³⁶ Laffont et al. (1998b) assume that receiver benefits are equal to zero, but—as noted above—this can be viewed as an extreme case of the constant-proportional-benefits assumption.

consumption benefits).³⁷ Berger (2004) assumes that receive prices fee must be set equal to 0 and shows that a carrier will tend to set its off-net send price above $c + a$ in order to make the rival network less attractive because subscribers to the latter will receive fewer incoming messages. In our model, the difference in on- and off-net pricing comes from the failure to internalize external benefits, not an affirmative attempt to harm the rival network.

The tendency for various models to lead to tipping underscores the importance of how one models consumer expectations with respect to ultimate network sizes. In models in which consumers form rational expectations of network sizes and carriers can commit to prices but not quantities, there may be multiple equilibria in which the market tips to a single carrier as end users flock to a carrier expected to be dominant. Such models also typically have degenerate equilibria in which all carriers fail precisely because end users expect them to fail (*i.e.*, the chicken-and-egg problem). In contrast, our Cournot model, in which carriers can commit to subscriber levels, tends to rule out extreme tipping and chicken-and-egg problems. Although such tipping and degenerate equilibria may be serious issues for new services, they are much less likely to be significant for established communications services, or for services where consumers have low switching costs (say due to number portability). The Cournot model can be viewed as a means of approximating a dynamic process in which consumer expectations with respect to network sizes change slowly over time because consumers observe network sizes and predict that these sizes will remain stable.

The Cournot assumption also has implications for carriers' incentives with respect to setting their off-network prices. Carrier i 's choice of off-net prices affects the consumer surplus

³⁷ Similarly, in a model of undifferentiated carriers, Degraha (2003, p. 224) finds that setting the access charge equal to the termination cost can lead to tariff-mediated network effects that induce tipping to a

that end users can expect to derive from subscribing to carrier j 's service. Under the homogeneous good, Cournot model, making carrier j 's service more or less desirable to consumers has no effect on carrier i 's profits—carrier j will simply adjust its hook-up to keep its subscriber level and, hence, consumer surplus, constant. Consequently, carrier i ignores the effects of its off-net prices on carrier j and its subscribers, and carrier i sets its off-net prices to maximize the sum of the benefits derived from off-net calling by carrier i and its subscribers.

This result need not hold under other models of carrier competition. As noted above, carrier i may benefit from setting its off-net prices to hurt carrier j when the carriers are Hotelling competitors. In the other direction, if the networks do not directly compete for customers, then carrier i might set its off-net prices to make carrier j more attractive, thus increasing the number of subscribers on network j and increasing the off-net calling benefits enjoyed by subscribers to network i . The importance of such cross-carrier effects is an empirical question that remains to be answered.

VII. CONCLUSION

In the absence of receiver benefits, the receiver is unwilling to pay to exchange messages, the sender of a message can be viewed as the “cost causer,” efficient pricing sets the send price equal to the marginal message cost, and the receiver’s network should recover its message costs from the sender or the sender’s carrier. The existence of receiver benefits fundamentally changes the analysis of interconnection charges. Instead of being how to recover the terminating network’s costs from the sender, the issue is how to recover the combined marginal costs of a

single carrier.

message between the sender and receiver in a way that efficiently internalizes the two-sided benefits.

As demonstrated by our formal analysis, this change in perspective has several broad implications:

- It is misleading to think of the originating carrier as purchasing termination services. One could just as well think of the terminating carrier as purchasing origination services. There are theoretical arguments for both positive and negative access charges.
- Second-best pricing generally entails both fixed and traffic-sensitive charges to both senders and receivers even when all costs are traffic sensitive.
- In important cases, imposition of an access charge can affect efficiency by altering the relative levels of the send and receive prices but cannot affect their sum, which remains inefficiently high.
- In absence of cross-network internalization mechanisms, such as repeat play or inter-dependent pricing strategies, off-net prices tend to be inefficient because efficient pricing by a service provider typically benefits the subscribers to rival providers and/or those rival providers.
- Even in highly symmetrical situations, an access charge of zero is not necessarily socially optimal.

VIII. APPENDIX

Proof of Proposition 1: We first prove several lemmas that lead to the proof of the main result.

Lemma A1: $p = 0 = r$ is not a solution to the welfare-maximizing pricing problem.

Proof: By assumption, $M(0) - 2c < 0$.³⁸ **QED**

Lemma A2: *If the hazard rate is everywhere increasing or everywhere decreasing, then any interior solution must be symmetric and there can be at most one such solution.*

Proof: The first-order conditions for an interior solution are $M(p) + r - 2c = 0$

and $M(r) + p - 2c = 0$. Setting these expressions equal to one another and rearranging terms, an interior solution must satisfy $M(p) - p = M(r) - r$, which can be written as

$$\frac{S(p)}{D(p)} = \frac{S(r)}{D(r)}.$$

By Lemma A1 of Hermalin and Katz (2004), if the hazard rate is strictly monotonic, then

$D(q)/S(q)$ is a strictly monotonic function of q , which implies $p = r$.

A symmetric interior solution must satisfy $M(p) + p = 2c$. The left-hand side of the equation is strictly increasing in p (with a derivative bounded below by 1), while the right-hand side is constant. By assumption, $M(0) < 2c$. Hence, $M(p) + p = 2c$ has exactly one solution.

QED

Lemma A3: *If the hazard rate is everywhere increasing, then there is a unique optimal price pair, (p^*, r^*) , and it satisfies $0 < p^* = r^* < c$.*

³⁸ Recall that 0 is the lower bound of the support of the distribution of message exchange values, so that $M(0)$ is the unconditional mean of ω , which is assumed to be less than $2c$.

Proof: Given the two lemmas above and the fact that efficient send and receive prices sum to less than $2c$, we need only show that there cannot be a corner solution with one of the two prices equal to 0. The proof proceeds by contradiction.

Suppose counterfactually that $p^* = 0$. Then it would have to be the case that

$$M(r^*) - 2c \geq 0 .$$

Having an interior solution with respect to r implies

$$M(0) + r^* - 2c = 0 .$$

These two expressions can be simultaneously satisfied only if

$$M(r^*) - r^* \geq M(0) - 0 .$$

But, applying the argument used in Lemma A2, this inequality cannot hold when the hazard rate is everywhere increasing. **QED**

Lemma A4: *If the hazard rate is everywhere decreasing, then both $p^* = 0, r^* = 2c - M(0)$ and $r^* = 0, p^* = 2c - M(0)$ are efficient price pairs.*

Proof: From the analysis above, we must show only that there cannot be a positive, symmetric solution. Suppose, counterfactually, that $p^* = q^* = r^*$ for some $q^* > 0$. Consider increasing one price by Δ and decreasing the other by Δ . The change in welfare is

$$\begin{aligned} & [S(q^* - \Delta)D'(q^* + \Delta) - S(q^* + \Delta)D'(q^* - \Delta)] \\ & + (2c - 2q^*)[D(q^* - \Delta)D'(q^* + \Delta) - D(q^* + \Delta)D'(q^* - \Delta)] . \end{aligned}$$

The first bracketed expression has the same sign as

$$\frac{D'(q^* + \Delta)}{S(q^* + \Delta)} - \frac{D'(q^* - \Delta)}{S(q^* - \Delta)} .$$

As shown in Hermalin and Katz (2004, proof of Proposition 2), a decreasing hazard rate implies that $D'(q)/S(q)$ is increasing in q . Hence, the first bracketed expression is positive. The second bracketed expression has the same sign as

$$\frac{D'(q^* + \Delta)}{D(q^* + \Delta)} - \frac{D'(q^* - \Delta)}{D(q^* - \Delta)} .$$

A decreasing hazard rate implies that this difference is positive (the hazard rate is $-D'(q)/D(q)$) and, hence, the second bracketed expression is also positive. Therefore, symmetric prices cannot be optimal. **QED**

Lastly, suppose that the hazard rate is constant. Then ω must have an exponential distribution, and $D(q) = e^{-q/\mu}$ for some constant, μ . As shown in Hermalin and Katz (2004), any two prices that sum to the same amount, say k , give rise to the same level of user surplus. They also give rise to the same quantity exchanged and, thus, the same production costs and carrier revenues.³⁹ It follows that total surplus depends solely on k . Setting $r = 0$ and $p = k$, the first-order condition derived above determines the optimal value of k : $M(0) + k - 2c = 0$.

Proposition 1 follows.

Proof of Proposition 2: By (4), n_i consumers will be willing to subscribe to network i only if

$$h_i = n_i (D(p_{ii})S(r_{ii}) + D(r_{ii})S(p_{ii})) + n_j (D(p_{ji})S(r_{ij}) + D(r_{ji})S(p_{ij})) - T(n_1 + n_2) .$$

We can decompose the problem of finding the prices that maximize a carrier's profits as:

$$\max_{p_{ii}, r_{ii}} D(p_{ii})S(r_{ii}) + D(r_{ii})S(p_{ii}) + (p_{ii} + r_{ii} - 2c)D(p_{ii})D(r_{ii}) \quad (15)$$

and

³⁹ The quantity exchanged is $D(p)D(k-p) = e^{-p/\mu} e^{-(k-p)/\mu} = e^{-k/\mu}$.

$$\max_{p_{ij}, r_{ij}} D(p_{ji})S(r_{ij}) + D(r_{ji})S(p_{ij}) + D(p_{ji})D(r_{ij})(r_{ij} - c + a) + D(r_{ji})D(p_{ij})(p_{ij} - c - a) . \quad (16)$$

Observe that (15) is equivalent to the corresponding social problem. The claim with respect to the off-net prices follows from differentiation of (16). **QED**

Proof of Proposition 3:

Necessity: The number of subscribers in a symmetric equilibrium must satisfy the weak inequality in (6) or the carriers would choose to shut down. By (5), any equilibrium must also satisfy the equality in (6).

Sufficiency: Suppose that N^d satisfies (6), and define

$$Z(\delta) \equiv \{N^d + 2\delta\}w^* + \frac{1}{2}N^d w^c + T(N^d + \delta) + \{\frac{1}{2}N^d + \delta\}T'(N^d + \delta) .$$

We know that $\frac{1}{2}N^d$ yields non-negative profits, so it is at least as good a response to $\frac{1}{2}N^d$ as is shutting down. By (5), a necessary condition for $\frac{1}{2}N^d + \delta > 0$ to be a best response to $\frac{1}{2}N^d$ is that $Z(\delta) = 0$. By construction, $Z(0) = 0$. Observe that

$$Z'(\delta) = 2\{w^* + T'(N^d + \delta)\} + \{\frac{1}{2}N^d + \delta\}T''(N^d + \delta) .$$

For any $\delta > 0$, $Z'(\delta) < 0$ by the inequality in (6) and the concavity of $T(\cdot)$. Hence, $Z(\delta) < 0$ for all $\delta > 0$.

Suppose, counterfactually, that there existed a $\delta_0 < 0$, such that $Z(\delta_0) = 0$ and this output yielded the carrier profits at least equal to those when $\delta = 0$:

$$Z(\delta_0) = \{\frac{1}{2}N^d + 2\delta_0\}w^* + \frac{1}{2}N^d (w^* + w^c) + T(N^d + \delta_0) + \{\frac{1}{2}N^d + \delta_0\}T'(N^d + \delta_0) = 0 \quad (17)$$

and

$$\{\frac{1}{2}N^d + \delta_0\}w^* + \frac{1}{2}N^d w^c + T(N^d + \delta_0) \geq 0 . \quad (18)$$

(17) and (18) imply

$$(\frac{1}{2}N^d + \delta_0)w^* + \{\frac{1}{2}N^d + \delta_0\}T'(N^d + \delta_0) \leq 0.$$

It follows that

$$Z'(\delta) = 2w^* + 2T'(N^d + \delta) + \{\frac{1}{2}N^d + \delta\}T''(N^d + \delta) < 0$$

for all $\delta \in (\delta_0, 0]$. Hence, it cannot be the case that $Z(\delta_0) = 0 = Z(\delta)$. Therefore, $\frac{1}{2}N^d$ must be the unique best response to $\frac{1}{2}N^d$.

Uniqueness of N^d : Suppose, counterfactually, that there are multiple symmetric equilibria, and let N_L and N_H denote the equilibrium number of subscribers in two such equilibria, where

$N_L < N_H$. We can rewrite the equality in necessary condition (6) as

$$\frac{1}{2}\{2w^* + w^c + T'(N^d)\}N^d + T(N^d) = 0$$

Taking the derivatives with respect to N yields

$$\frac{1}{2}\{2w^* + w^c + T'(N)\} + T'(N) + \frac{1}{2}N^dT''(N) .$$

By (6) and the concavity of $T(\cdot)$, we know that $0 \geq w^* + T'(N) \geq \frac{1}{2}(w^* + w^c) + T'(N)$ for all $N \geq N_L$. Hence, the derivative above is negative over the interval $[N_L, N_H]$, and the necessary condition cannot be satisfied by both values of N . **QED**

Proof of Corollary to Proposition 3: First, suppose $T(0) > 0$ and $w^* + T'(0) \leq 0$. As $N \downarrow 0$,

$$\frac{1}{2}\{w^* + w^c\}N + T(N) \rightarrow T(0) > 0 \text{ and } -\frac{1}{2}N\{w^* + T'(N)\} \rightarrow 0 . \text{ As } N \uparrow \bar{N} ,$$

$$\frac{1}{2}\{w^* + w^c\}N + T(N) \rightarrow -\infty \text{ and } -\frac{1}{2}N\{w^* + T'(N)\} \rightarrow \infty . \text{ It follows that there exists at least one}$$

$$N_0 > 0 \text{ such that } \frac{1}{2}\{w^* + w^c\}N_0 + T(N_0) + \frac{1}{2}N_0\{w^* + T'(N_0)\} = 0 . \text{ } w^* + T'(0) \leq 0 \text{ implies that}$$

$$w^* + T'(N_0) < 0 . \text{ Hence, it must be the case that } \frac{1}{2}\{w^* + w^c\}N_0 + T(N_0) > 0 . \text{ } N_0 \text{ thus satisfies}$$

the necessary and sufficient conditions in the hypothesis of Proposition 3.

Next, suppose that $T(0) \leq 0$ and $\frac{1}{2}(w^* + w^c) + T'(0) \leq 0$. By the concavity of $T(\cdot)$, $0 \geq \frac{1}{2}\{w^* + w^c\}N + NT'(0) + T(0) > \frac{1}{2}\{w^* + w^c\}N + T(N)$ for all $N > 0$ and, by Proposition 3, there is no symmetric equilibrium with positive sales. **QED**

Proof of Proposition 6: To see why it can never be socially optimal to set the access charge to induce an asymmetric interior equilibrium, observe that, by Proposition 5, when such equilibria exist there is a family of equilibria all entailing the same industry subscription level, \hat{N} . Hence, the corner equilibria yield the greater total surplus because all calls are made at the efficient prices and the resulting level of total surplus is $\hat{N}w^* + \int_0^{\hat{N}} T(s)ds$.⁴⁰ **QED**

Proof of Proposition 7: The result with respect to existence and equilibrium network sizes can be seen almost immediately from inspection of condition (6) and application of Proposition 3. Label the corresponding industry subscription levels N_L and N_H , with $N_L < N_H$.

Equilibrium profits are $\pi^d(N, w^c) \equiv \frac{1}{2}N^2(w^* + w^c) + NT(N)$. As shown by Proposition 3, in equilibrium $\pi^d(N, w^c) = -\frac{1}{2}N^2\{w^* + T'(N)\}$. Differentiation of the right-hand side of this equation yields $-N\{w^* + T'(N) + \frac{1}{2}NT''(N)\}$. $T''(N) < 0$ by assumption. By (6), $w^* + T'(N_L) \leq 0$. Hence, $-\frac{1}{2}N^2\{w^* + T'(N)\}$ is an increasing function of N over the interval $[N_L, N_H]$.

⁴⁰ The proof assumes that $\bar{w}^c \leq w^*$ holds as a strict inequality. Even if it held as an equality, there would be no loss in social welfare from restricting the choice of outcome to the monopoly and symmetric duopoly equilibria.

Equilibrium consumer surplus is equal to $\int_0^N T(s)ds - NT(N)$. Differentiation yields

$-NT'(N) > 0$. **QED**

Proof of Proposition 8: Recall that the necessary conditions for the existence of a monopoly equilibrium are that there exists an N^m such that

$$-\frac{T(N^m)}{N^m} \geq w^c \quad (19)$$

and

$$-\frac{T(N^m)}{N^m} - T'(N^m) = 2w^* \quad (20)$$

By the concavity of $T(\cdot)$ and $T(0) \geq 0$, $0 > T(0) - T(N) + NT'(N) \geq -T(N) + NT'(N)$. Hence,

$$-T'(N^m) > -\frac{T(N^m)}{N^m} \quad (21)$$

(20) and (21) imply that

$$-\frac{T(N^m)}{N^m} < w^* \quad (22)$$

By (6), a necessary condition for the symmetric duopoly equilibrium when $w^c = \bar{w}^c$ is

$$-\frac{T(N^d)}{N^d} - T'(N^d) = 2w^* + \bar{w}^c + \frac{T(N^d)}{N^d} \quad (23)$$

Observe that the derivative of $-\frac{T(N)}{N}$ is $-\frac{NT'(N) - T(N)}{N^2} \geq -\frac{NT'(N) + T(0) - T(N)}{N^2} > 0$,

⁴¹ It follows that, for \bar{w}^c sufficiently close to w^* , there is no monopoly equilibrium when the access charge is set at the level that induces \bar{w}^c .

where the first inequality follows from $T(0) \geq 0$ the second inequality follows from the concavity of $T(\cdot)$. Thus, the left-hand sides of both (20) and (23) are increasing in N , while the right-hand side of (23) is decreasing in N . It follows that the sign of $N^d - N^m$ is equal to the sign of

$$\bar{w}^c + \frac{T(N^d)}{N^d} . \quad (24)$$

$T(N)/N$ is decreasing in N . Hence, for \bar{w}^c sufficiently close to w^* , (22) implies that (24) is positive when $N^d - N^m \leq 0$. $N^d - N^m \leq 0$ would thus contradict the fact about signs just derived. *Reductio ad absurdum*, $N^d > N^m$, when \bar{w}^c is sufficiently close to w^* .

The levels of total surplus under the monopoly and duopoly outcomes are

$$W^m \equiv (N^m)^2 w^* + \int_0^{N^m} T(s) ds$$

and

$$W^d \equiv (N^d)^2 \frac{1}{2}(w^* + w^c) + \int_0^{N^d} T(s) ds = (N^d)^2 (w^* - \delta) + \int_0^{N^d} T(s) ds ,$$

respectively, where $\delta \equiv \frac{1}{2}(w^* - w^c)$. Observe that

$$\int_{N^m}^{N^d} T(s) ds > \{N^d - N^m\} T(N^d) \geq -\{N^d - N^m\} w^* N^d ,$$

where the second inequality follows from (6). Hence,

$$W^d - W^m > w^* \{N^d - N^m\} N^m - (N^d)^2 \delta . \quad (25)$$

As $\delta \downarrow 0$, N^m remains unchanged and N^d increases but is subject to a finite upper bound.

Hence, for all δ sufficiently close to 0, the right-hand side of (25) is positive and total surplus is greater under the duopoly outcome than the monopoly one. It is evident that consumer surplus is also greater under the duopoly outcome.

For sufficiently low values of \bar{w}^c , (19) will be satisfied and a monopoly equilibrium will exist even when $w^c = \bar{w}^c$. The fact that (19) is satisfied implies that (24) would be negative were $N^d - N^m > 0$, which would contradict the fact derived above. It follows, that $N^d \leq N^m$ for sufficiently low values of \bar{w}^c . In this case, the symmetric duopoly equilibrium is clearly inferior—it garners weakly few subscribers and there is inefficient off-net message exchange.⁴² **QED**

⁴² More generally, this argument implies that, if both the monopoly and symmetric duopoly outcomes are equilibria for a given value of w^c , then the monopoly outcome is welfare superior to the duopoly outcome that arises at that value of w^c .

REFERENCES

- Armstrong, Mark, "Network Interconnection in Telecommunications," *The Economic Journal*, Vol. 108, No. 448. (May, 1998), pp. 545-564.
- Armstrong, Mark, "The Theory of Access Pricing and Interconnection," in *Handbook of Telecommunications Economics*, Vol. 1, Martin E. Cave, Sumit K. Mujumdar, and Ingo Vogelsang, (Eds.), North-Holland: Amsterdam, 2002.
- Atkinson, Jay M. and Christopher C. Barnekov, "A Competitively Neutral Approach to Network Interconnection," December 2000. OPP Working Paper Series #34, Federal Communications Commission.
- Berger, Ulrich, "Access Charges in the Presence of Call Externalities," *Contributions to Economic Analysis & Policy*, Vol. 3, No. 1, Article 21 (2004).
- Bolt, Wilko and Alexander F. Tieman, "Social Welfare and Cost Recovery in Two-sided Markets," Unpublished manuscript, Research Division, De Nederlandsche Bank, 2005.
- DeGraba, Patrick, "Efficient Intercarrier Compensation for Competing Networks when Customers Share the Value of a Call," *Journal of Economics & Management Strategy*, Vol. 12, No. 2 (Summer, 2003), pp. 207-230.
- European Commission, "Commission Recommendation of [...] on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU," undated draft, *available at* http://ec.europa.eu/information_society/policy/ecomm/doc/library/public_consult/termination_rates/termination.pdf, site visited 14 April 2009.
- Farrell, Joseph, and Garth Saloner, "Standardization and variety," *Economics Letters*, Vol. 20 (1986), pp. 71-74.
- Federal Communications Commission, *Developing a Unified Intercarrier Compensation Regime*, CC Docket No. 01-92, Further Notice of Proposed Rulemaking, FCC 05-33 (released March 3, 2005).
- Hermalin, Benjamin E., and Michael L. Katz, "Network Interconnection with Two-Sided User Benefits," unpublished manuscript, University of California, Berkeley, 2001a.
- _____, "Retail Telecommunications Pricing in the Presence of External Effects," in Gary Madden, ed., *International Handbook on Emerging Telecommunications Networks*, Cheltenham, UK: Edward Elgar Publishing, 2001b.
- _____, "Sender or Receiver: Who Should Pay to Exchange an Electronic Message?" *The RAND Journal of Economics*, Vol. 35 (2004), pp. 423-448.
- _____, "Intercarrier Compensation with All-You-Can-Eat Retail Pricing," unpublished manuscript, University of California, Berkeley, 2006.

- Jeon, Doh-Shin, Jean-Jacques Laffont, and Jean Tirole, "On the 'Receiver-Pays' Principle," *The RAND Journal of Economics*, Vol. 35 (2004), pp. 85-110.
- Katz, Michael L., "Observable Contracts as Commitments: Interdependent Contracts and Moral Hazard," *Journal of Economics & Management Strategy*, Vol. 15 (2006), pp. 685-706.
- Katz, Michael L., and Carl Shapiro, "Network externalities, Competition, and Compatibility," *American Economic Review*, Vol. 75 (1985), pp. 424-440
- Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole, "Internet Interconnection and the Off-Net-Cost Pricing Principle," *The RAND Journal of Economics*, Vol. 34 (2003), pp. 370-390.
- Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole, "Network Competition: I. Overview and Nondiscriminatory Pricing" *The RAND Journal of Economics*, Vol. 29, No. 1. (Spring, 1998a), pp. 1-37.
- Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole, "Network Competition: II. Price Discrimination" *The RAND Journal of Economics*, Vol. 29, No. 1. (Spring, 1998b), pp. 38-56.
- Laffont, Jean-Jacques, and Jean Tirole, *Competition in Telecommunications*, Cambridge : MIT Press, 2000.
- Taylor, Lester D., *Telecommunications Demand in Theory and Practice*, Dordrecht: Kluwer Academic Publishers, 1994.
- Valletti, Tommaso M., and Carlo Cambini, "Investments and Network Competition," *The RAND Journal of Economics*, Vol. 36 (2005), pp.446-467.
- Willig, Robert, "The Theory of Network Access Pricing," in *Issues in Public Utility Regulation*, ed. by H. Trebing. Michigan State University Press, East Lansing, MI. 1979.