Network Interconnection with Two-Sided User Benefits

Benjamin E. Hermalin & Michael L. Katz
Walter A. Haas School of Business
University of California, Berkeley

July 2001

Abstract

Previous work on network interconnection has tended to overlook that *both* the sender and receiver of an electronic message take actions, bear costs, and derive benefits from the message exchange. In a simple model with two-sided benefits and fixed network architectures, we find that the socially optimal interconnection charge is independent of the "direction" of the message and is used to induce optimal end-user prices for sending and receiving messages that account for demand conditions. These optimal retail prices depend solely on the *sum* of the marginal costs of exchanging a message across the two networks, not the specific marginal costs of the individual networks. Optimal interconnection pricing with endogenous network investment is also explored.

Contents

Acknowledgements

The authors wish to acknowledge the support of the Walter A. Haas School of Business through the Willis H. Booth and Edward & Mollie Arnold Chairs and the University of California through a Committee on Research Grant. The views expressed here are solely those of the authors. As such, they do not necessarily reflect the views of the chair donors, the University of California, or consulting clients in government and the telecommunications industry.

Contact Information

By mail:

University of California
Walter A. Haas School of Business
545 Student Services Building #1900
Berkeley, CA 94720–1900
USA

By e-mail: `hermalin@haas.berkeley.edu` or `katz@haas.berkeley.edu`.

1 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 anothers' networks. Interconnection can thus significantly affect efficiency and market structure. Consequently, interconnection rules, particularly those concerning intercarrier compensation, are one of the most important — and contentious — areas of public policy concerning telecommunications markets.[1]

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] With a few notable exceptions, previous theoretical work on interconnection pricing has ignored the benefits enjoyed by the receiving party.[3] This treatment typically is justified by one of two assumptions: either the receiving party enjoys no benefits from a message exchange or the effects between two parties are internalized. The first assumption clearly is unrealistic. If it were correct, we would never answer the telephone or read our e-mail. The second assumption is applicable only to certain situations in which either the two communicating parties are altruistic with respect to one another or have a repeated relationship.[4]

Although our focus is on inter-carrier pricing, the welfare consequences of interconnec-

[1]For recent statement of many of the issues, see Federal Communications Commission (2001).

[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 of the no-receiver-benefits case include Armstrong (1998), Laffont and Tirole (2000), and Laffont et al. (1998).

[4]Willig (1979, pages 124–25) establishes conditions under which call externalities will be internalized in the demand for sending messages and thus can be incorporated into the standard analysis of access externalities. Essentially the condition is that sending a message triggers a set number of incoming messages. Hermalin and Katz (2001b, §3.5) develop a simple game-theoretic model in which users can (partially) internalize call externalities by engaging in tit-for-tat message initiation.

tion charges derive in large part from their effects on the resulting retail prices set by the networks. In discussions of telecommunications pricing, it is often asserted that the principle of cost-causative pricing implies that it is efficient for the sender to pay the marginal cost of exchanging a message. In the absence of receiver benefits, this claim is correct: The sender can be viewed as the "cost causer," efficient pricing is purely cost based, and efficient pricing sets the send price equal to the marginal message cost. 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. Indeed, when the receiver derives no benefits from the exchange, he is unwilling to pay for messages.

The presence of receiver benefits changes all of these findings. First, one could just as well assert that the receiver causes the costs by accepting the message. Second, efficient prices must internalize the external effects across the two parties to a message exchange. Hence, efficient pricing requires consideration of demand conditions, as well as cost conditions. Third, it frequently is not socially optimal to have one party bear the full marginal costs of exchanging a message. Given the significant effects that receiver benefits have on optimal retail prices, it is not surprising that receiver benefits can also have strong effects on optimal intercarrier-charges.

Atkinson and Barnekov (2000) and DeGraba (2000a,b, 2001) recently proposed interconnection pricing regimes based on analyses recognizing that the parties at both ends of a message enjoy benefits. DeGraba (2000a) investigates socially optimal interconnection charges in a model in which the sender and receiver enjoy equal benefits from any given message exchange and the networks to which the sender and receiver subscribe have equal marginal costs. DeGraba argues that having an interconnection charge of 0 — so-called *bill and keep* — is efficient. As DeGraba notes, one needs to consider the implications of a broader set of distributions for sender and receiver valuations and of unequal network costs. Below, we provide that analysis and demonstrate that non-zero interconnection charges may be optimal in the presence of receiver benefits.

Atkinson and Barnekov (2000) examine a model in which parties at both ends benefit from message exchange, and they conclude that each carrier should recover from its own subscribers all costs not incremental to interconnection itself. Critically, Atkinson and Barnekov assume that retail prices are independent of the interconnection pricing regime. However, interconnection costs represent either marginal message costs or fixed per-subscriber costs. In either event, these costs will affect a carrier's profit-maximizing retail prices. In contrast to Atkinson and Barnekov, we allow retail prices to reflect the interconnection charge. In fact, in our basic model, the sole role of interconnection charges is to influence retail prices. Not surprisingly, we reach very different conclusions than do Atkinson and Barnekov.

There is a related literature on transfers between firms that are part of a payment system (*e.g.,* a bank that handles the account of a merchant that accepts a credit card and the bank that issues that credit card). This work can be reinterpreted in terms of telecommunications carrier interconnection charges.[5] Because both merchants and consumers may value card use, this literature examined the implications of two-sided benefits. In an early analysis, Baxter (1983) identified the use of the interconnection charge as a means to internalize what would otherwise be external effects between the end users. In a more recent analysis, Schmalensee (2001) implicitly solves for welfare-optimal interconnection charge for the case of linear demands by both sides of the market, where he allows for both monopoly and oligopoly among the "carriers."[6]

We characterize socially optimal inter-carrier charges for the interconnection of two networks. In Section 3, we assume that the interconnecting carriers do not compete with each other. This situation arises, for example, when a mobile telephone user calls a landline local exchange carrier's customer. Alternatively, the sender's network could be a long-distance carrier completing a call on the local exchange network to which the receiver subscribes. We

[5]One must be careful about extending application of the theory in the opposite direction — there are important effects that arise because card-accepting merchants may compete with one another and may have commercial relationships with the other sides of transactions.

[6]His analysis focuses on privately optimal interconnection charges.

consider both bilateral monopoly and markets in which there is competition at each end of the message (*e.g.,* competing long-distance carriers). Our analysis demonstrates that non-zero interconnection charges often are optimal, and that the role of interconnection charges is to share the costs of the message exchange efficiently between the two users. The optimal retail prices depend on demand conditions and the *sum* of marginal costs, not the individual networks' costs. However, the optimal interconnection charge does depend on the networks' individual costs because those costs affect the carriers' choices of retail prices. In stark contrast with the no-receiver-benefits view of the world, the socially optimal interconnection charge does not depend on the direction of the message exchange.

In Section 4, we examine markets in which the interconnecting carriers compete with one another. These carriers can be thought of as competing local exchange telephone service providers or competing Internet backbone service providers.[7] We demonstrate how the optimal interconnection charge depends on both demand conditions and the costs of the higher-cost carrier, even when it carries no traffic in equilibrium. Moreover, in those cases where both network have positive equilibrium traffic volumes, the traffic balance between them is irrelevant.

The analysis proceeds as follows. Section 2 introduces notation and preliminary concepts. Our basic model is used to examine interconnection between non-competing carriers in Section 3 and competing carriers in Section 4. Each of the next two sections then relaxes one of the major assumptions of the basic model. Section 5 allows for greater stochastic dependence between the sender and receiver's message values, while Section 6 allows for endogenous network architectures and costs. The paper closes with a brief conclusion.

[7] In a paper written independently of ours, Laffont et al. (2001) also examine this case. The relationship between the two analyses is discussed below.

2 Preliminaries

2.1 Model Structure

Consider two individuals, A and B, who may wish to communicate. We model the communication between A and B as the exchange of a single *message*, which can be a telephone call, a paging message, an SMS message, a data file, or an e-mail, for example. 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 communication 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 in this paper is restricted to one-way calling models.[8]

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.

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

[8]Previous theoretical examinations of retail pricing in the presence of call externalities, Hahn (2000), Squire (1973), and Srinagesh and Gong (1996), implicitly examine the one-way calling case. Hermalin and Katz (2001c) examine retail pricing by a single network in both one-way and two-way calling models.

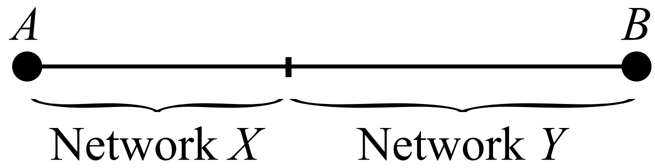

Figure 1: Schematic of message transit between consumers A and B across X and Y's networks.

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 again be approximated by a one-way calling model in which a party sends a message whenever her expected value of message exchanges exceeds the price she must pay.[9]

Initially, there are two networks in our model, with user A connected to network X, and user B connected to network Y. Figure 1 illustrates. The cost of exchanging a message between A and B is $m = m_X + m_Y$, where m_X and m_Y are the non-negative marginal costs incurred by networks X and Y, respectively.

To simplify the analysis of message exchange, we assume the set of parties connected to the networks is invariant with respect to pricing. Alternatively, consumers can be thought of as subscribing to an always-on service, and our model can be interpreted as examining the probability with which someone connects to a network.[10] Under either interpretation, an external effect arises because the receiver takes an action that effects whether the sender can exchange a message with him. We also abstract from repeated-play considerations: each party is motivated only by his or her private net benefits for a given potential message exchange.

We ignore income effects by assuming that consumer j derives gross monetary benefits v_j

[9]More precisely, we have a pair of one-way calling models — see Hermalin and Katz, 2001c for an analysis of a pair of one-way calling models in the context of a single network provider.

[10]For analyses of network connection decisions, see Hahn (2000), Littlechild (1975), Rohlfs (1974), and Squire (1973).

from communication. These benefits should be interpreted as net of any opportunity costs incurred in exchanging messages (*e.g.,* answering a telemarketing call during dinner).[11] The parties' values of communicating, v_A and v_B, could be unknown to them at the time they make their send and receive decisions. Specifically, each individual has some prior knowledge (type, signal, etc.), $\omega_j \in \Omega_j$. The pair (ω_A, ω_B) has joint distribution $\Psi(\omega_A, \omega_B)$, which is common knowledge. Each (ω_A, ω_B) vector defines a joint distribution over (v_A, v_B).

Observe that under the information structure presented above, either party may draw inferences about his or her value of message exchange from the other party's behavior. The possibility of complex inferences makes the analysis of a general model difficult.[12] Progress can be made if one assumes each party's information is relevant only for predicting his or her own value of communicating because then neither party draws on the behavior of the other to form inferences about his or her own value of exchanging a message. To that end, we make the following assumption throughout our analysis:

Assumption 1 *For* $i = A, B$, $\Omega_i \subseteq \mathbb{R}$ *and* $v_i = \omega_i + \eta_i$, *where* η_i *is a random variable satisfying* $\mathbb{E}\{\eta_i|\omega_i\} = \mathbb{E}\{\eta_i|\omega_i, \omega_j\} \equiv 0$.[13]

Under this assumption, ω_i contains no additional information useful to j in predicting the expected value of v_j *conditional* on the information he or she already possesses, ω_j. Observe, however, that, *ex ante*, the parties' information, ω_i and ω_j, could be correlated. Observe, too, that ω_i is now party i's expected value of communicating.

---

[11] These opportunity costs could be sufficiently large that it could be efficient to charge positive message prices even when $m = 0$, either to discourage message exchange (*e.g.,* a tax on "junk" email) or to raise revenues to subsidize the other party to engage in message exchange. See Hermalin and Katz (2001b) for a discussion.

[12] In a companion paper (Hermalin and Katz, 2001a), we examine the effects of such inferences more fully. Hermalin and Katz (2001b) also contains examples drawn from the general model.

[13] Given the parties' behavior, what is important about their information is how it maps to their expected values of calling. That is, one could assume that each party's information lies in some multi-dimensional set and that there is some mapping from this set to expected values. The set Ω_i would then be the range of this mapping for party i. What we are assuming in Assumption 1 is that this mapping not change when a party learns *new* information from observing the behavior of the other party (*e.g.,* that he is called).

Assumption 1 strikes us as reasonable when applied to the *sender* because the receiver's decision whether to accept often will not convey much information to the sender about the value to her of having sent the message. This relationship is particularly likely in markets where there are multiple potential senders and the receiver does not know the identity of the party sending the message at the time the acceptance decision is made. Moreover, in this latter situation, application of Assumption 1 to the *receiver* is plausible: It may be unreasonably complex for the receiver to draw inferences about the message's value to him from the fact that a message was sent. When the receiver knows who is sending him a message (*e.g.,* the receiver has caller ID or can see the sender's email address), the assumption that the receiver infers nothing from being sent a message could be less innocuous. On the other hand, if the sender is known to send messages that vary greatly in value to the receiver, or if the receiver's message value strongly depends on his current opportunity cost of time (*e.g.,* whether he's eating dinner or not), then the receiver's estimation of a message's value to him may be little influenced by the sender's having initiated the message.

Our basic model also assumes that ω_A and ω_B are independently distributed, with distribution functions $\Psi_A(\cdot)$ and $\Psi_B(\cdot)$, respectively.[14] Let $S_i(\omega) \equiv 1 - \Psi_i(\omega)$ be the *survival function*. If $\psi_i(\cdot)$ is the density associated with $\Psi_i(\cdot)$, then $S_i'(\omega) = -\psi_i(\omega)$. Thoughout our basic model, we assume:

Assumption 2 *For $i = A, B$, there exists finite $\underline{\omega}_i$ and $\bar{\omega}_i$ such that $S_i(\underline{\omega}_i) = 1$, $S_i(\bar{\omega}_i) = 0$, and $S_i(\cdot)$ is twice differentiable, with $S_i'(\cdot) < 0$ and $S_i''(\cdot) \leq 0$, on $\left[\underline{\omega}_i, \bar{\omega}_i\right]$.*

Examples of concave survival functions include those derived from convex power-function distributions. Power-function distributions describe random variables, U, defined on an interval $[u_1, u_2]$, such that $\Pr\{U \leq u\} = \left(\frac{u-u_1}{u_2-u_1}\right)^\kappa$. If $\kappa \geq 1$, the distribution is everywhere weakly convex (note $\kappa = 1$ is the uniform distribution).

[14]The independence of ω_A and ω_B does not imply that v_A and v_B are independently distributed; η_A and η_B may still be correlated.

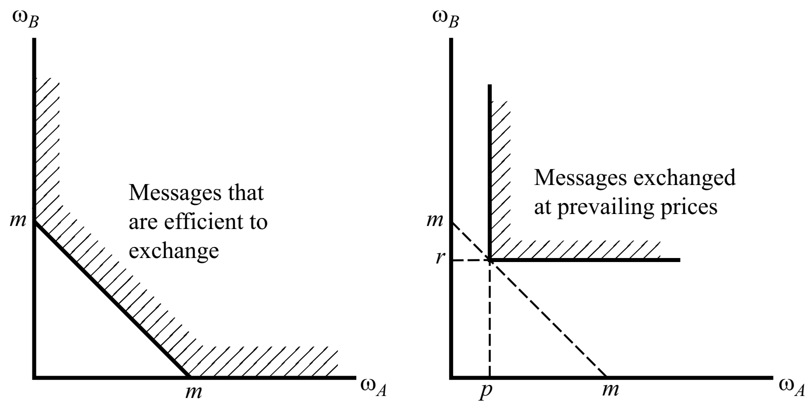

Figure 2: Messages that are efficient to exchange and those that are exchanged at prices p and r.

We take total surplus, the sum of producer and consumer surplus, as our welfare measure. Under the *first-best outcome,* total surplus is maximized if all messages for which $v_A+v_B > m$ are exchanged and no messages for which $v_A + v_B < m$ are exchanged. Because v_A and v_B might be realized only after a message is sent and received, the first-best outcome is generally unattainable. A more realistic welfare standard is second-best, or *information-constrained*, efficiency: A message is exchanged if and only if the social expected value conditional on what the parties know exceeds the cost. As the left panel of Figure 2 shows, it is information-constrained efficient to exchange all messages above the line $\omega_A + \omega_B = m$.[15]

Throughout, we assume that, with positive probability, both A and B expect message exchange to be valuable:

Assumption 3 $S_A\big(\lambda m\big)S_B\big((1-\lambda)m\big) > 0$ *for some* λ*.*

[15]For convenience, we have drawn the figures for cases in which all messages have non-negative values to both parties. Nothing in our analysis precludes messages with negative values.

2.2 Two Retail-Pricing Benchmarks

Although our focus is on inter-carrier pricing, the welfare consequences of interconnection charges in our basic model derive from charges' effects on the carriers' resulting retail prices. Hence, we begin by characterizing the first-best retail prices and then identifying two useful retail pricing benchmarks.

When both parties benefit from a message exchange, it is feasible to charge both the sender and receiver positive prices. Examples of positive receive prices in the U.S. include collect calling, 800 numbers, and most wireless telephone service. Let p denote the per-message price paid by the sender, and let r denote the per-message price paid by the receiver. These prices are paid if and only if the message is successfully exchanged.

Given the retail prices, a consumer equilibrium comprises a pair of mutual best-response functions: a sending rule for A and an acceptance rule for B. Under Assumption 1, a weakly dominant strategy for A is to send a message if and only if $\omega_A \geq p$. Likewise a weakly dominant strategy for B is to accept the message if and only if $\omega_B \geq r$. Because it relates the probability of purchase to the price user i faces, the survival function can be viewed as user i's demand function. The right panel of Figure 2 shows the set of messages exchanged when users face prices p and r. As a comparison of the two panels shows, even if below-cost pricing is feasible, it generally is impossible to achieve efficient message exchange using only simple send and receive prices.

We now consider two retail price benchmarks. The first is Ramsey (1927) prices — the prices that maximize total surplus subject to the constraint that they cover the message costs. Ramsey prices must satisfy the constraint that that $p + r \geq m$.[16] For later use, it is helpful to consider a modified version of the Ramsey problem. Let π denote a non-negative constant. We are interested in characterizing socially optimal prices subject to the constraint that $p + r = m + \pi$.

[16]To focus on the effects of receiver benefits on pricing, we assume there are no network fixed costs.

Given prices p and $r = m + \pi - p$, expected welfare, $\mathbb{E}W$, is

$$\mathbb{E}W = \int_{m+\pi-p}^{\infty} \left(\int_{p}^{\infty} \big(\omega_A + \omega_B - m\big) \psi_A(\omega_A) d\omega_A \right) \psi_B(\omega_B) d\omega_B \,.$$

Recalling the definition of $S_i(\cdot)$ and integrating by parts,

$$\mathbb{E}W = S_B(m + \pi - p) \int_{p}^{\infty} S_A(\omega) d\omega + S_A(p) \int_{m+\pi-p}^{\infty} S_B(\omega) d\omega + \pi S_A(p) S_B(m + \pi - p) \,. \quad (1)$$

The expression for expected welfare, equation (1), is stated solely in terms of m, the sum of the individual networks' marginal costs.[17] Intuitively, there is no sense in which the costs of a given network are associated with a specific user, as opposed to a specific sender-receiver pair. Hence, the welfare question is how to share the sum of the costs efficiently between the two users:

Observation 1 *The optimal send and receive prices depend solely on the* sum *the networks' marginal costs, $m_X + m_Y$, and not the individual components.*

As should be clear, this result holds under much broader conditions than those of our basic model.[18]

The following result is a straightforward generalization of Proposition 2 in Hermalin and Katz (2001c), and its proof is omitted here.

Proposition 1 *Suppose that ω_A and ω_B are independently and identically distributed according to the survival function $S(\cdot)$. Then, for any non-negative constant π, the prices $p = r = (m + \pi)/2$ maximize total surplus subject to the constraint that $p + r = m + \pi$.*

[17] Although our maintained interpretation of m is that it is the sum of the networks' marginal message costs, in Propositions 1 and 2, which follow, m can also be interpreted as social marginal cost (*e.g.,* including congestion costs imposed on other users).

[18] This result was first identified by Srinagesh and Gong (1996). One situation in which it does not hold is when a user can choose among alternative networks that have different marginal costs and give rise to different distributions of benefits between the two users. See footnote 25 below.

The next proposition (proved in the Appendix) extends a second result of Hermalin and Katz (2001c) to characterize optimal retail prices in certain asymmetric cases, including survival functions derived from power function distributions.

Proposition 2 *Suppose that ω_A and ω_B are independently distributed according to the survival functions $S_A(\cdot)$ and $S_B(\cdot)$, respectively. If $\psi_j(\omega)$ crosses $\psi_i(\omega)$ once from above at $\hat{\omega} \geq \frac{m+\pi}{2}$,[19] then party i pays more than party j under any socially optimal pricing scheme that satisfies the constraint $p + r = m + \pi$, $\pi \geq 0$.*

Another useful benchmark is provided by a profit-maximizing, monopoly network. It is helpful to divide the monopolist's problem into two steps. First, suppose the monopolist's profit margin is fixed at π per message exchanged: $p + r - m = \pi$. The profit maximizer chooses a price pair from this line that maximizes the probability of message exchange:

$$S_A(p)S_B(r) = S_A(p)S_B(\pi + m - p)\,. \tag{2}$$

The corresponding first-order condition for an interior maximum is[20]

$$S_A'(p)S_B(\pi + m - p) - S_A(p)S_B'(\pi + m - p) = 0\,. \tag{3}$$

When $S_A(\cdot) \equiv S(\cdot) \equiv S_B(\cdot)$, $p = \frac{1}{2}(\pi + m)$ is a solution, which implies $r = \frac{1}{2}(\pi + m)$. Equation (3) is sufficient, as well as necessary, under our assumption that $S(\cdot)$ is concave.

Once the optimal prices p and r for a given π are found, the second step is to maximize expected profits with respect to π. When $S_A(\cdot) \equiv S(\cdot) \equiv S_B(\cdot)$, the profit maximizer's problem is to

$$\max_{\pi} S\left(\frac{\pi + m}{2}\right) S\left(\frac{\pi + m}{2}\right) \pi$$

[19] That is, $\psi_j(\omega) > \psi_i(\omega)$ if $\omega < \hat{\omega}$ and $\psi_j(\omega) < \psi_i(\omega)$ if $\omega > \hat{\omega}$.

[20] Here, and throughout our analysis, we can limit attention to interior maxima. Given Assumption 2, only interior equilibria exist as long as a carrier does not wish to shut down. By Assumption 3, it is not privately optimal for the monopoly network to shut down. Similarly, below it will never be in the interest of policy makers to set the interconnection charge at a level that would lead to an equilibrium with no message exchange.

Because the derivative of this expression with respect to π is positive at $\pi = 0$ (it equals S^2), the usual monopoly distortion of too few sales holds. Note, however, that conditional on this distortion, the profit-maximizing monopolist imposes no further distortion in terms of the optimal division of the cost and markup between sender and receiver. When $S_A(\cdot)$ is not identical to $S_B(\cdot)$, the profit-maximizing monopolist may distort the relative levels of the send and receive prices, in addition to setting the markup up too high. The relative-price distortion arises because — for a given margin — the profit maximizer chooses the relative prices solely to maximize the probability that a message will be exchanged, while the welfare maximizer also takes into account the value of the messages exchanged.

3 Interconnection of Non-Competing Carriers

This section and the following one present the results for our basic model. In each case, we make two broad assumptions that are relaxed in later sections. First, we assume that the A and B's expected message values are independently distributed. Second, we take the carriers' network architectures as exogenous. The difference between the models of the two sections lies in the assumption about whether the interconnecting carriers compete with one another. In the present section, we assume that they do not. Within this framework, we consider two market structures. In one, each carrier has a monopoly with respect to one of the subscribers. The sender's only choice of carrier is network X, and the receiver's only choice of carrier is network Y. This situation corresponds to one in which A subscribes to a monopoly provider of mobile telephone or long distance telephone service, while B subscribes to a monopoly landline local exchange provider. Later in the section, we allow for competition between multiple carriers at a given end of the message.

Let t denote the interconnection or access charge paid by network X to network Y when a message is carried from A to B. At this point, t may be positive or negative.

Network X chooses p to maximize:

$$S_A(p)S_B(r)(p - m_X - t)\,. \qquad (4)$$

The corresponding first-order condition for an interior maximum is

$$S'_A(p)(p - m_X - t) + S_A(p) = 0\,. \tag{5}$$

Similarly, network Y chooses r to maximize:

$$S_A(p)S_B(r)(r - m_Y + t)\,, \tag{6}$$

and the corresponding first-order condition for an interior maximum is

$$S'_B(r)(r - m_Y + t) + S_B(r) = 0\,. \tag{7}$$

Assumption 2 ensures that (5) and (7) are sufficient as well as necessary.

Because of the multiplicative structure of demand, each network sets a price that depends solely on its costs (including the effects of any interconnection charge) and the distribution of its subscribers' message exchange values. In particular, a change in r has no effect on X's profit-maximizing choice of p and vice versa. Each network sets its margin equal to one over the hazard rate for its subscriber's valuation. Let $p_b(q) \equiv \arg\max_p S_A(p)(p - q)$ and $r_b(q) \equiv \arg\max_r S_B(r)(r - q)$ denote the maximizing price for networks X and Y, respectively.

As expected, the usual Cournot (1838) complements problem arises, and the bilateral monopolists set retail send and receive prices that sum to more than those that would be set by a single monopoly network:

Proposition 3 *For any interconnection charge, t, bilateral monopoly networks set higher prices than would a single monopoly network with cost $m_X + m_Y$.*

Proof: Adopt the shorthand $p_b = p_b(m_X + t)$ and $r_b = r_b(m_Y - t)$. ¿From (5) and (7):

$$p_b - m_X - t = \frac{-S_A(p_b)}{S'_A(p_b)} > 0 \text{ and } r_b - m_Y + t = \frac{-S_B(r_b)}{S'_B(r_b)} > 0\,. \tag{8}$$

Hence,

$$S'_A(p_b)\big(p_b + r_b - (m_X + m_Y)\big) + S_A(p_b) < 0 \tag{9}$$

and

$$S'_B(r_b)\big(p_b + r_b - (m_X + m_Y)\big) + S_B(r_b) < 0\,. \tag{10}$$

Multiplying the left-hand sides of (9) and (10) by $S_B(r)$ and $S_A(p)$, respectively, we get the derivatives of the *single* monopoly network's profits with respect to p and r, respectively. Because, as shown, these derivatives are negative when evaluated at p_b and r_b, we can conclude that p_b and r_b are each set at a higher level than a single monopoly network would set them. (It is readily shown that Assumption 2 is sufficient for the single monopolist's profit-maximization problem to be globally concave.) ■

Now consider the choice of t. The analysis of the socially optimal inteconnection charge turns on two straightforward, but important, observations. First, the interconnection charge affects relative prices: an increase in t generally raises the equilibrium value of p and lowers the equilibrium value of r. Second, the level of t may affect the equilibrium level of the sum of the retail prices, $p + r$.

The next lemma is an immediate consequence of the fact that the two networks face identical optimization problems when $S_A \equiv S_B$ and $t = (m_Y - m_X)/2$.

Lemma 1 *Suppose that ω_A and ω_B are independently and identically distributed according to the survival function $S(\cdot)$. If $t = (m_Y - m_X)/2$, then there exists a unique equilibrium and $p = r$.*

Combined with Proposition 1, Lemma 1 tells us that $t = (m_Y - m_X)/2$ induces efficient retail prices *conditional* on the industry margin. The next result characterizes t's effect on that margin. We present the proof, which relies on a standard monopoly comparative statics result, in the Appendix.

Lemma 2 *Suppose that ω_A and ω_B are independently and identically distributed according to the survival function $S(\cdot)$. If $S''' \geq 0$, then $t = (m_Y - m_X)/2$ minimizes the equilibrium industry margin, $p + r - m$.*

These two lemmas establish conditions under which $t = (m_Y - m_X)/2$ both induces the surplus-maximizing retail prices among those having a constant industry margin and induces the lowest equilibrium industry margin. Combining these results, one obtains

Proposition 4 *Suppose that ω_A and ω_B are independently and identically distributed according to the survival function $S(\cdot)$. If $S''' \geq 0$, then the socially optimal interconnection charge for bilateral monopolists is $t = (m_Y - m_X)/2$.*

As do all of our propositions, Proposition 4 builds on the assumption that $S''(\cdot) \leq 0$. If one relaxes this assumption, one can obtain a striking result: When $S(z) = e^{-z/\nu}$, the level of t is irrelevant. Simple calculations show that any pair of retail prices such that $p+r = m+\pi$ yield expected welfare $(2\nu+\pi)e^{-(m+\pi)/\nu}$. Moreover, the equilibrium prices are $p = \nu+m_X+t$ and $r = \nu + m_X - t$, so that the industry margin is independent of t. Hence, for *any* value of t, expected welfare is $4\nu e^{-(2+m/\nu)}$.

Proposition 4 indicates that the socially optimal interconnection charge can depend on the values of the individual networks' marginal costs. It is important to recognize, however, that t depends on the distribution of the total marginal cost between the two networks only because those costs affect the networks' retail pricing decisions; the socially optimal retail prices themselves are independent of the distribution of costs between the networks.

That t is driven by demand considerations can be seen by considering the following example in which the two parties' message values are not identically distributed. Suppose that ω_A is uniformly distributed on $[0, 1]$, while ω_B is uniformly distributed on $[0, \beta]$, β a positive constant. Then $S_A(p) = 1 - p$, and $S_B(r) = 1 - r/\beta$; uniform distributions give rise to linear demand functions. Simple calculations demonstrate that $p = (1 + m_X + t)/2$ and $r = (\beta+m_Y-t)/2$. The equilibrium industry margin is independent of t: $p+r = (1+\beta-m)/2$.

Hence, when $S''(q) \equiv 0$, t is set solely to induce socially optimal prices given the margin. Somewhat more tedious calculations demonstrate that expected welfare is

$$\begin{aligned} &\frac{1}{\beta}[\beta - r][1-p]\left(\frac{1}{2}[\beta + r + 1 + p] - m\right) \\ &= \left(\frac{\beta - m_Y + t}{2}\right)\left(\frac{1 - m_X - t}{2}\right)\Lambda\,, \end{aligned}$$

where $\Lambda = (\frac{1}{2}[\beta + r + 1 + p] - m)$ is a constant with respect to t. The welfare optimal interconnection charge is consequently

$$t = \frac{m_Y - m_X + 1 - \beta}{2}\,. \tag{11}$$

Therefore, if $m_X = m_Y$, the optimal interconnection charge is negative if $\beta > 1$ and positive if $\beta < 1$.

Applying Proposition 2 to this example, observe that $\psi_i(\omega) = \min\left\{\frac{1}{\beta}, 1\right\}$ and $\psi_j(\omega) = \max\left\{\frac{1}{\beta}, 1\right\}$. If

$$0 \leq \frac{m}{4} + \min\{1, \beta\} - \frac{1}{2}\max\{1, \beta\}\,,$$

then ψ_j crosses ψ_i once from above at or to the right of $\frac{1}{2}(m+\pi)$. Proposition 2 then implies t should be chosen to induce $p > r$ if $\beta < 1$ and induce $p < r$ if $\beta > 1$. The value of t given by equation (11) satisfies this condition.

In practice, many carriers face competition. Suppose that there are two networks at each end of a potential message exchange, X_1 and X_2 which compete to serve user A, and Y_1 and Y_2 which compete to serve user B. The carriers at each end are undifferentiated Bertrand (1883) competitors with one another. However, the interconnecting carriers (*e.g.,* X_2 and Y_1) do not compete with each other.

By the usual analysis, it is straightforward to establish:[21]

[21] Kaplan and Wettstein (2000) establish that the only Nash equilibrium is the pure-strategy one when, as here, $S(q)q$ is bounded as $q \to +\infty$ (see Assumption 2).

Lemma 3 *Suppose that at each end the higher-cost carrier's marginal cost is less than the lower-cost carrier's monopoly price. The unique Bertrand equilibrium prices given* t *are* $p = \max\{m_{X_1}, m_{X_2}\} + t$ *and* $r = \max\{m_{Y_1}, m_{Y_2}\} - t$*.*

As is typical in Bertrand models, if the carriers have unequal costs marginal costs, the lower-cost supplier prices "just below" the marginal cost of the other one. With Bertrand competition, the equilibrium margin is unaffected by the interconnection charge regardless of the shape of the survival functions. Moreover, the value of t does not affect network competition — the lower-cost network serves all subscribers. The sole effect of the interconnection charge is to induce different relative retail prices. The socially optimal interconnection charge is the one that induces Ramsey prices that recover $\max\{m_{X_1}, m_{X_2}\} + \max\{m_{Y_1}, m_{Y_2}\}$ in an efficient way from the two users. It follows from Proposition 1 that

Proposition 5 *Suppose that* ω_A *and* ω_B *are independently and identically distributed according to the survival function* $S(\cdot)$*. If there is Bertrand competition at each end of the message, then the socially optimal interconnection charge is*

$$t = \frac{\max\{m_{Y1}, m_{Y2}\} - \max\{m_{X_1}, m_{X_2}\}}{2}.$$

Similarly, it follows from Proposition 2 that

Proposition 6 *Suppose that* ω_A *and* ω_B *are independently distributed according to the distribution functions with densities* $\psi_A(\cdot)$ *and* $\psi_B(\cdot)$*, respectively, where* $\psi_j(\omega)$ *crosses* $\psi_i(\omega)$ *once from above at*

$$\hat{\omega} \geq \frac{1}{2}\left(\max\{m_{X_1}, m_{X_2}\} + \max\{m_{Y_1}, m_{Y_2}\}\right).$$

With competing carriers at each end, the socially optimal inteconnection charge is greater than $(\max\{m_{Y_1}, m_{Y_2}\} - \max\{m_{X_1}, m_{X_2}\})/2$ *if* $i = A$ *and less than that if* $i = B$*.*

One can also consider mixed cases. For example, there might be competition among wireless or long-distance providers to originate a call, but a *monopoly* local exchange carrier terminates it. In this case, $p = \max\{m_{X_1}, m_{X_2}\} + t$ while the value of r depends on whether the local exchange carrier is regulated. Suppose it is not. Then $r = r_b(m_Y - t)$, and the optimal interconnection charge depends on two effects. First, it is readily shown that, when $S_A(\cdot) \equiv S_B(\cdot)$, to induce $p = r$ requires $t > (m_Y - \max\{m_{X_1}, m_{X_2}\})/2$ in order to correct for the differential markup strategies pursued by the two networks carrying the traffic in equilibrium. Second, in addition to affecting the relative levels of p and r, the interconnection charge affects the industry margin when $r_b'(m_Y + t) \neq 1$. The optimal t depends on the sum of these two effects. Now suppose the local exchange carrier were regulated and forced to set $r = m_Y - t + \sigma$, for some $\sigma \geq 0$. Then the industry margin would be independent of t. Hence, when $S_A \equiv S_B$, the socially optimal interconnection charge induces $p = r$, or $t = (m_Y + \sigma - \max\{m_{X_1}, m_{X_2}\})/2$. The effects of other forms of regulation remain to be explored.

It is clear from inspection that, if one reverses the values of S_A and S_B and of m_X and m_Y in any of the analyses above, the sign of the optimal interconnection charge is reversed as well. Hence, another important observation is the following.

Observation 2 *The socially optimal interconnection charge is independent of the calling direction.*

The analysis thus far has assumed that a given carrier can bill only its own subscribers. We close this section by considering what happens when monopoly carriers at either end of a message exchange can directly levy charges on their own customer and the other carrier's customer. That is, carrier $Z = X, Y$ sets send and receive prices p_Z and r_Z, respectively. The following result is proved in the Appendix.

Proposition 7 *When each bilateral monopolist can bill both ends of a message exchange, there is a unique equilibrium level of message exchange and it is independent of the inter-*

connection charge. Moreover, in equilibrium, the two carriers earn equal margins,

$$p_X + r_X - m_X - t = p_Y + r_Y - m_Y + t\,.$$

When each carrier has the authority to bill the other carrier's customers, policy makers cannot meaningfully affect retail prices through the interconnection charge. For a given industry margin, such carriers act like an integrated monopolist to maximize the probability of message exchange without regard for the values of different exchanges.[22] Thus, in symmetric situations, the power to bill the other network's customers does not have adverse welfare effects on relative prices because equal send and receive prices are socially optimal and will be chosen by the carriers. But in general, this type of billing undermines the ability of policy makers to use interconnection charges to induce socially optimal retail prices.

4 Interconnection of Competing Carriers

The model of the previous section assumed that the interconnecting carriers did not compete with one another. In this section, we consider the case in which the two interconnecting networks are undifferentiated Bertrand competitors with one another.[23]

Suppose that there are two networks, either of which can compete for either type of user. Continue to assume that user A originates messages and user B receives them so that the two carriers compete for A's patronage in terms of a send price and B's patronage in terms of a receive price. By the usual Bertrand analysis, it is again straightforward to establish:

---

[22]This property can be seen from the first-order necessary conditions for p_Z and r_Z presented in the proof of Proposition 7 in the Appendix. Inspection shows that $S_A'(p_X + p_Y)S_B(r_X + r_Y)$ must equal $S_A(p_X + p_Y)S_B'(r_X + r_Y)$ in equilibrium, which is the first-order condition for maximizing $S_A(p_X + p_Y)S_B(r_X + r_Y)$ subject to the constraint that $p_X + p_Y + r_X + r_Y$ is equal to some constant.

[23]Laffont et al. (2001) also examine optimal interconnection charges for the case of competing carriers. Unlike the present section, they also allow for differentiated networks and they examine the effects of payments between the sender and receiver. They do not consider the extensions of this case that we examine in Sections 5 and 6 below.

Lemma 4 *Suppose that the higher-cost carrier's marginal costs are less than the lower-cost carrier's monopoly price. Given t, the unique Bertrand equilibrium prices are $p_Z = m_Z + t$ and $r_Z = m_Z - t$, where $m_Z = \max\{m_X, m_Y\}$.*

Note that these are the equilibrium prices even when one carrier wins both types of subscriber and does not actually pay the interconnection fee.[24]

Like the case of Bertrand competition at each end of a message, which we examined in the previous section, Lemma 4 indicates that both the equilibrium margin and the ability of either carrier to win sales in equilibrium are unaffected by the level of the interconnection charge. Again, the sole effect of the interconnection charge is to induce different relative retail prices. There is, however, an important difference between the two cases. When the interconnecting firms are competitors with one another, the lower-cost network serves all subscribers. Hence, the interconnection charge does not have to correct for the possibility that the sender's network has different marginal costs than the receiver's network. Instead, the socially optimal interconnection charge is the one that induces Ramsey prices that recover $2\max\{m_X, m_Y\}$ in an efficient way from the two users.

Applying Proposition 1, one obtains the following result.

Proposition 8 *Suppose that ω_A and ω_B are independently and identically distributed according to the survival function $S(\cdot)$ and that the higher-cost carrier's marginal cost is less than the lower-cost carrier's monopoly price. Then bill and keep is the socially optimal interconnection regime with competing interconnected carriers (*i.e., $t = 0$ *maximizes equilibrium total surplus subject to the constraint that $p + r = 2\max\{m_X, m_Y\}$).*

Comparing the findings for the cases of competition at each end with competition between the interconnecting carriers (Propositions 5 and 8, respectively), one sees that the socially

[24] Laffont et al. (2001) call this the *off-net-cost pricing principle.* A similar result arises in the intellectual property licensing literature (see, for example, Katz and Shapiro, 1985), where an intellectual property owner acts as if it is paying itself a royalty because displacing the output of rivals has an opportunity cost equal to the foregone royalty. Laffont et al. provide an extensive and insightful analysis of the conditions under which the result holds.

optimal interconnection charge accounts for cost differences in the first case but not in the second. This is a consequence of the fact that competition between the interconnecting carriers ensures that the lower-cost network serves both ends of a message exchange and there are no cost differences for which to correct.

Asymmetric cases can be addressed by applying Proposition 2 to obtain the following result.

Proposition 9 *Suppose that* ω_A *and* ω_B *are independently distributed according to the survival functions* $S_A(\cdot)$ *and* $S_B(\cdot)$*, respectively, and* $\psi_j(\omega)$ *crosses* $\psi_i(\omega)$ *once from above at* $\hat{\omega} \geq \max\{m_X, m_Y\}$*. With competing interconnected carriers, the socially optimal interconnection charge is positive if* $i = A$ *and negative if* $i = B$*.*

If the two networks have unequal costs, only the lower-cost network carries traffic in equilibrium. If $m_X = m_Y$, both carriers can serve end users in equilibrium, and can do so in any proportion or mix. None of the analysis of the socially optimal interconnection charge depends on the mix of customers each network obtains. In other words,

Observation 3 *The socially optimal interconnection charge is independent of the traffic balance.*

5 Stochastically Dependent Message Values

The basic model assumes that the sender and receiver's expected message values are independently distributed. We now allow ω_A and ω_B to be perfectly correlated. Specifically, let $\omega_B \equiv \alpha + \beta\omega_A$, where $\beta > 0$. Figure 3 illustrates.

We first characterize the Ramsey optimal retail prices. As can readily be seen from Figure 3, the information-constrained optimal prices are

$$p^* = \frac{m + \pi - \alpha}{1 + \beta} \text{ and } r^* = \frac{\beta(m + \pi) + \alpha}{1 + \beta}\,. \tag{12}$$

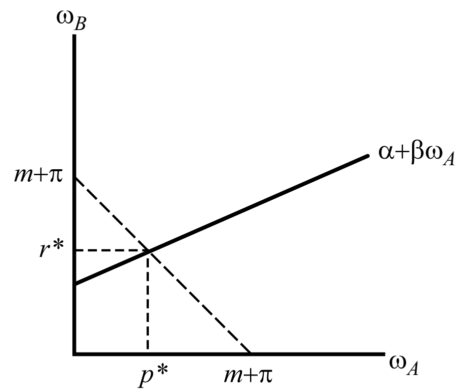

Figure 3: The correlated values case.

Observe that these prices do not depend on any other properties of the joint distribution of ω_A and ω_B. In particular, it is possible with positive correlation that the party who faces the higher price has the lower expected value of message exchange (*e.g.*, $\alpha + \beta\omega$ crosses ω from above, $\frac{\alpha}{1-\beta} > m/2$, and most of the weight is on $\omega_A \geq \frac{\alpha}{1-\beta}$).

We first examine the socially optimal interconnection charge when networks X and Y are bilateral monopolists as in Section 3. Consider X's profit-maximizing choice of p given Y's choice of r. Given r, X sets p to maximize

$$\text{Prob}[\omega_A \geq p \text{ and } \omega_B \geq r](p - t - m_X) = \text{Prob}\left[\omega_A \geq \max\left\{p, \frac{r-\alpha}{\beta}\right\}\right](p - t - m_X).$$

Network X does better to set p equal to $(r-\alpha)/\beta$ rather than less than $(r-\alpha)/\beta$ because the former response yields the same message volume as the latter, but at a higher profit per message. Hence, $p \geq (r-\alpha)/\beta$ and, thus, $\text{Prob}[\omega_A \geq p \text{ and } \omega_B \geq r] = S_A(p)$. The latter equation implies that X's expected profits are $S_A(p)(p - m_X - t)$. Recall that $p_b(q) \equiv \arg\max_p S_A(q)(p-q)$. Because $S_A(p)(p - m_X - t)$ is concave in p, X sets p at the maximum of $(r-\alpha)/\beta$ and $p_b(m_X + t)$. Summarizing X's reaction function:

$$p^e(r) = \begin{cases} p_b(m_X + t) & \text{if } r \leq \alpha + \beta p_b(m_X + t) \\ \frac{r-\alpha}{\beta} & \text{otherwise} \end{cases}.$$

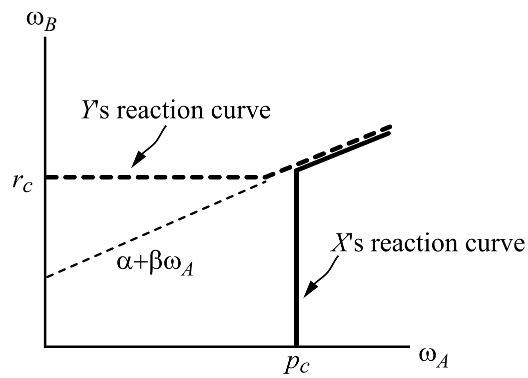

Figure 4: Reaction functions for X (solid line) and Y (dashed line). The set of equilibria are where both reaction functions coincide with the $\omega_B = \alpha + \beta\omega_A$ line.

Note, for low r, X chooses the same price as it would were ω_B distributed independently of ω_A. A similar analysis for Y reveals that

$$r^e(p) = \begin{cases} r_b(m_Y - t) \text{ if } p \leq \frac{r_b(m_Y - t) - \alpha}{\beta} \\ \alpha + \beta p \text{ otherwise} \end{cases},$$

where, as before, $r_b(m_Y - t) \equiv \arg\max_r S_B(r)(r - m_Y + t)$.

Figure 4 illustrates the reaction curves for both carriers simultaneously. As the figure shows, there is a continuum of equilibria of the form $p = q$ and $r = \alpha + \beta q$, with the lowest value of q being

$$\max\left\{p_b(m_X + t), \frac{r_b(m_Y - t) - \alpha}{\beta}\right\}.$$

For a given t, the equilibrium with the lowest value of q is socially optimal. This conclusion follows from the fact that welfare is a decreasing function of q for all $q \geq \frac{m-\alpha}{1+\beta}$. Moreover, for a given t, both networks prefer this equilibrium to any other. This conclusion follows from the fact that $p \geq p_b(m_X + t)$ and $r \geq r_b(m_Y - t)$ in all of the equilibria and each network's profits are a concave function of its own price and a non-increasing function of the other network's price. We will focus on this equilibrium in our analysis of the socially optimal interconnection charge.

The welfare-optimal interconnection charge minimizes

$$\max\left\{p_b(m_X+t), \frac{r_b(m_Y-t)-\alpha}{\beta}\right\}.$$

Observe that $p_b(m_X+t)$ is increasing in t, while $r_b(m_Y-t)$ is decreasing. Hence, if there exists a value of t for which $\alpha+\beta p_b(m_X+t)=r_b(m_Y-t)$, then that value of t is welfare maximizing. As the next proposition shows, such a value of t exists.

Proposition 10 *Suppose that $\omega_B \equiv \alpha+\beta\omega_A$, where $\beta>0$. The socially optimal interconnection charge under bilateral monopoly is*

$$t=\frac{m_Y-\beta m_X-\alpha}{1+\beta}.$$

Proof: By hypothesis,

$$S_A\left(\frac{q-\alpha}{\beta}\right)=S_B(q).$$

Hence, p_b and r_b solve

$$\max_p S_A(p)(p-m_X-t)$$

and

$$\max_r S_A\left(\frac{r-\alpha}{\beta}\right)(r-m_Y+t),$$

respectively. The first-order conditions are

$$S_A'(p)(p-m_X-t)+S_A(p)=0$$

and

$$S_A'\left(\frac{r-\alpha}{\beta}\right)\frac{1}{\beta}(r-m_Y+t)+S_A\left(\frac{r-\alpha}{\beta}\right)=0,$$

respectively. Substituting in

$$t=\frac{m_Y-\beta m_X-\alpha}{1+\beta}$$

and rearranging algebraically, these equations become

$$S_A'(p)\left(p-\frac{m_X+m_Y-\alpha}{1+\beta}\right)+S_A(p)=0$$

and

$$S_A'\left(\frac{r-\alpha}{\beta}\right)\left(\frac{r-\alpha}{\beta}-\frac{m_X+m_Y-\alpha}{1+\beta}\right)+S_A\left(\frac{r-\alpha}{\beta}\right)=0\,,$$

respectively. By Assumption 2, the first-order conditions are sufficient as well as necessary and, moreover, $S_A'(q)(q-k)+S_A(q)=0$ has a unique solution for any k. Hence,

$$\frac{r_b-\alpha}{\beta}=p_b\,.$$

■

DeGraba (2000a) analyzed the case of $\alpha=0$, $\beta=1$, and $m_X=m_Y$, and found that $t=0$ is socially optimal. The present analysis shows extends this finding to alternative values of α, β, and unequal network marginal costs, and it shows that non-zero values of t are socially optimal outside of the special case considered by DeGraba.

To conclude this section, suppose that X and Y are Bertrand competitors. Recall that when the higher-cost carrier's marginal costs are less than the lower-cost carrier's monopoly price, the unique Bertrand equilibrium prices are $p_Z(t)=m_Z+t$ and $r_Z(t)=m_Z-t$, where $m_Z=\max\{m_X,m_Y\}$. Suppose that $m_X\leq m_Y$. Then the equilibrium margin is $2(m_Y-m_X)$. The interconnection charge affects only the relative send and receive prices, not the equilibrium margin. The socially optimal interconnection charge satisfies $p_Z(t)=p^*$ and $r_Z(t)=r^*$. The first of these equalities holds if

$$m_Y+t=\frac{2m_Y-\alpha}{1+\beta}$$

(where the right-hand side follows by substituting $2m_X$ for m and $2(m_Y-m_X)$ for π in (12)). Consequently,

$$t=\frac{(1-\beta)m_Y-\alpha}{1+\beta}\,.$$

It is readily checked that this value of t also equates $r_Z(t)$ and r^*. With independent and identically distributed message values, the socially optimal interconnection charge was always 0. With perfectly positively correlated values, nonzero interconnection charges can be optimal, although identical expected values for A and B (*i.e.,* $\beta = 1$ and $\alpha = 0$) again yield bill and keep as the optimal interconnection regime.

At first glance, the dependence of the optimal value of t on m_Y is surprising because, in equilibrium, no traffic is carried by network Y. The higher-cost carrier's marginal costs matter solely because they affect the equilibrium margin earned by the other carrier. The larger is m_Y, the larger is the sum of the equilibrium values of p and r. When $\beta \neq 1$, the interconnection charge has to adjust to ensure the efficient relative levels of p and r.

6 Endogenous Network Architectures

To this point, we have taken the networks' costs to be exogenously given. In practice, carriers may choose among alternative technologies or invest in reducing their costs. Interconnecting carriers also make choices with respect to where the their networks meet. These choices may be affected by the interconnection charge, and thus incentives effects on network architecture choices should be taken into account in setting socially optimal interconnection charges.

In this section, we explore these issues under the assumption that ω_A and ω_B are independently and *identically* distributed with common survival function $S(\cdot)$. Recall that, by Propositions 4 and 8, the socially optimal interconnection fee with exogenous network costs is $t = (m_Y - m_X)/2$ for bilateral monopolists (when $S'''(\cdot) \geq 0$) and 0 for competing interconnected carriers.

Suppose the carrier can make expenditures in R&D or equipment that lower its marginal costs. With bilateral monopolists, a fall in one network's marginal costs raises the other carrier's profits if t is held fixed. For example, Y earns

$$S(p_b(m_X + t))S(r)(r - m_Y + t)\,,$$

which is an decreasing function of m_X. Consumers also gain from a reduction in costs because it is partially passed through in the form of lower retail prices. Hence, a network's private incentives for cost reduction are lower than the social incentives. The underinvestment problem is magnified if the interconnection charge is adjusted to the socially optimal level that reflects the marginal cost reduction. Because $t = (m_Y - m_X)/2$, a reduction in a network's marginal costs shifts the interconnection charge in the direction that is unfavorable to that network. A similar disincentive effect would arise if regulation forced a network to lower its retail prices by a large fraction of its marginal cost reduction.[25] To internalize the investment incentives, the interconnection fee would have to be adjusted in the direction that favored the network that lowered its marginal costs. Starting from an initially optimal interconnection charge, this change would result in the other network's subscribers bearing an inefficiently high proportion of the marginal message costs. Hence, there is a tradeoff between inducing dynamic efficiency and allocative efficiency.

With competing interconnected carriers, the social benefits of cost reductions by the lower-cost carrier fully accrue to that carrier, as long as they are not drastic (*i.e.,* the carrier's post-reduction monopoly price is not less than its rival's marginal costs). However, cost reductions in other circumstances lead to private benefits that are lower than the social ones. For instance, suppose the initially higher-cost carrier invests in cost reduction but remains the higher-cost carrier. It still makes no sales in equilibrium, and thus realizes no private benefits. However, total surplus rises because the other carrier is forced to set the sum of its send and receive prices closer to the efficient level.

Interestingly, the level of the interconnection charge can indirectly affect cost-reduction

[25] A similar issue arises with consumer technology choice. When A gets a constant share of the net benefits under any technology, she chooses the technology that maximizes total surplus. Suppose, however, that a more costly technology generates additional benefits for A but not for B (*e.g.,* the technology also allows A to consume other services, such as cable television, that do not involve B.) In this case, a pricing scheme that sets r higher for the more expensive technology in order to cover the higher marginal message cost will tend to bias A's choice toward that technology. In effect, B would be subsidizing A's consumption of the additional benefits associated with the more costly technology. (DeGraba (2000a) provides an intuitive discussion of this issue.)

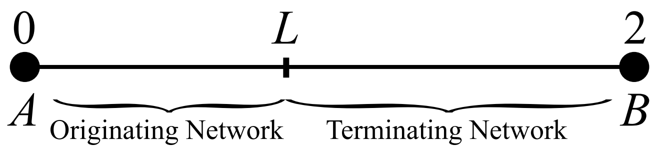

Figure 5: Endogenous meeting of originating and terminating networks at location L.

incentives. Suppose that a carrier can reduce is marginal costs by $\delta(I)$ by investing I. For non-drastic innovations, the lower-cost carrier's incentives to reduce its marginal costs further by δ are $\delta S(p)S(r)$.[26] Suppose that X is the lower-cost network. Recall that competition leads to $p_X + r_X = 2m_Y$. Thus, the value of t that maximizes X's investment incentives is the one that maximizes $S(p)S(r)$ subject to the constraint that $p_X + r_X = 2m_Y$. However, as we saw in the monopoly analysis, for any given margin, the carrier's profits are maximized by maximizing $S(p)S(r)$. Therefore, setting t to induce the profit-maximizing prices (given the margin) maximizes the lower-cost carrier's cost-reduction incentives. (Recall, this value of t would typically be non-zero for the case $S_A(\cdot) \neq S_B(\cdot)$ because the competitive prices, $p_X = m_Y + t$ and $r_X = m_Y - t$, do not directly reflect the different demand conditions of A and B.)

Now, consider the second dimension of the problem: the choice of meet point or place of interconnection. Suppose that A is located along a line at 0 and B located at 2, and that it costs either network μ per-unit distance to carry traffic. Suppose that A subscribes to network X and B subscribes to network Y. Let L denote the location along the line where one network hands off inter-network traffic to the other. Then $m_X = \mu L$ and $m_Y = \mu(2-L)$ for a message that traverses both networks. Figure 5 illustrates the situation. Note that if one network serves both parties and keeps the message entirely on its network, it incurs cost $m_Z = 2\mu$.

When the interconnecting carriers are bilateral monopolists, the optimal interconnection

[26] A non-drastic innovation is one such that the innovator's monopoly price remains above its rival's marginal cost.

charge is $t = \mu(1-L)$. Thus, if the two carriers have to meet each other halfway, the socially optimal interconnection charge is 0, but otherwise it could be positive or negative.

Where will networks exchange traffic? While a general analysis is beyond the scope of this paper, there is an interesting case that arises with telephone service. A vexing policy question has been what determines whether an entity is an end user, who pays retail prices, or a carrier, which participates in the interconnection regime. If entities are allowed to self designate or become carriers after meeting minimal requirements, and if the interconnection charge is set at zero, then user A could "interconnect" with, say, network Y at point 0. A would incur no transport costs and Y would incur costs of 2μ. In this case, it might appear that $t = \mu$ would be socially optimal. However, A would "sell" the service to himself at cost, while Y would mark up the charge to B. The different mark-up rules raise the same issues as in the mixed cases discussed after Proposition 6.

Lastly, consider competing interconnected carriers. Suppose that user A subscribes to network X and the carriers are bidding for the subscription of user B. If X wins the bidding at price r_X, it incurs costs of 2μ to carry the traffic end-to-end, and it earns $p_X + r_X - 2\mu$. If Y wins the bidding, network X earns $p_X - \mu L - t$. Hence, X is willing to set r_X as low as $(2-L)\mu - t$ to win B's patronage. If Y loses the bidding for B's patronage, Y earns nothing, while, if it wins, it earns $r_Y - (2-L)\mu + t$. Hence, like X, network Y is willing to price as low as $(2-L)\mu - t$ to win B's patronage, and this is the equilibrium receive price. Similar calculations show that the equilibrium send prices are $p_X = p_Y = L\mu + t$.[27] Hence, the socially optimal interconnection charge is again $t = \mu(1-L)$.

Again, the socially optimal interconnection charge depends on L. Here, an interesting case arises on the Internet, where major backbone service providers currently exchange traffic at an interconnection charge of zero, a practice known in the industry as settlement-free peering. Internet backbone carriers typically engage in "hot-potato routing," under which the originating network dumps traffic onto the other network as soon as it can. Hence, when

[27] We are assuming that X and Y also interconnect at L when X serves B and Y serves A.

traffic flows from A to B, one would expect L to be small and a positive interconnection charge to be socially optimal. Importantly, a positive interconnection fee is socially optimal not because A is the "cost causer," but because A's network has the opportunity to dump traffic onto the other network after a short distance. The ability to dump traffic resides with the network originating any given packet. Hence, when packets are exchanged in both directions as part of an Internet session, the originating network is determined on a per-packet basis rather than simply being the network whose customer initiated the session.[28]

7 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 is purely cost based, and efficient pricing sets the send price equal to the marginal message cost. Thus, 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. The analysis above demonstrates how the existence of receiver benefits fundamentally changes the analysis of interconnection charges. The receiver's carrier can charge its subscriber, and efficient retail prices must internalize the external effects across the two parties to a message exchange. The issue is no longer one of recovering the terminating network's costs from the sender. Instead, the issue is how to apportion the combined marginal costs of a message between the sender and receiver in an efficient way, accounting for pricing distortions due to carrier market power and possible effects on carrier investment incentives. While the cost recovery perspective suggests that $t = m_Y$ is optimal, our analysis found that, in the presence of positive receiver benefits, t would have this value only by coincidence. We also found that $t = 0$ could be efficient.

[28] Pairs of major Internet backbones typically interconnect with one another at several geographically diverse points. Whichever carrier whose user was originating a given packet would follow hot-potato routing and dump that packet on the other network at the interconnection point closest to the originating user.

Our analysis is based on simplifying assumptions. While the lessons drawn from our simple model will remain, further insights into optimal pricing could be gained by examining more general cases in which sending and receiving values are stochastically dependent, two-way calling is feasible, there is competition among differentiated carriers, and message exchange involves more than two networks (as is the case with long-distance telephone calls in the U.S., which typically involve an interexchange carrier and two local exchange carriers).

Appendix

Proof of Proposition 2: Let $i = A$ and $j = B$ (the proof when these are reversed is a straightforward variant of the one presented here).

> **Lemma:** *Distribution Ψ_A dominates distribution Ψ_B in the sense of first-order stochastic dominance ($\Psi_A \underset{\mathsf{FSD}}{\geq} \Psi_B$).*
>
> **Proof:** Define $Z(\omega) \equiv \Psi_A(\omega) - \Psi_B(\omega)$ We need to establish that $Z(\omega) \leq 0$ for all ω. Because $\psi_A(\omega) < \psi_B(\omega)$ for all $\omega < \hat{\omega}$, $Z(\omega) < 0$ for all $\omega \leq \hat{\omega}$. Recall $\bar{\omega} = \max \operatorname{supp}\{\omega_A\}$. $Z(\omega)$ is increasing in ω for $\omega \in (\hat{\omega}, \bar{\omega}_A)$ and $Z(\bar{\omega}_A) = 0$. Consequently, $Z(\omega) \leq 0$ for $\omega \in (\hat{\omega}, \bar{\omega}_A)$. ■

Observe that $S_B \equiv S_A + Z$. Substituting $S_A + Z$ for S_B in expression (1) yields the following expression for expected welfare:

$$\begin{aligned} &S_A(m+\pi-p)\int_p^\infty S_A(\omega)d\omega + S_A(p)\int_{m+\pi-p}^\infty S_A(\omega)d\omega + \pi S_A(p)S_A(m+\pi-p) \\ &+ Z(m+\pi-p)\int_p^\infty S_A(\omega)d\omega + S_A(p)\int_{m+\pi-p}^\infty Z(\omega)d\omega + \pi S_A(p)Z(m+\pi-p)\,. \end{aligned} \tag{13}$$

Define $V(p)$ to equal the first line of (13). Differentiating (13) with respect to p yields:

$$V'(p) - Z'(m+\pi-p)\left(\pi S_A(p) + \int_p^\infty S_A(\omega)d\omega\right) + S_A'(p)\left(\pi Z(m+\pi-p) + \int_{m-p}^\infty Z(\omega)d\omega\right) \tag{14}$$

Consider $p \in \left[m + \pi - \hat{\omega}, \frac{m+\pi}{2}\right]$. From Proposition 1, $p = \frac{m+\pi}{2}$ would maximize $V(p)$; hence, the first term of (14) is non-negative because $p \leq \frac{m+\pi}{2}$. From $m + \pi - p \leq \hat{\omega}$, it follows that $-Z'(m + \pi - p) \geq 0$ and, thus, the middle term is non-negative. The third term is the product of two negative quantities and, thus, positive as well. Therefore, (14) is positive if $p \in \left[m + \pi - \hat{\omega}, \frac{m+\pi}{2}\right]$. Because social welfare is weakly concave in p (recall Assumption 2), (14) must also be positive for $p < m + \pi - \hat{\omega}$ as well, which means (14) can be zero only for $p > \frac{m+\pi}{2}$. That is, A must pay a higher price than B. ■

Proof of Lemma 2: Each firm solves a problem of the form $\max S(q)(q - k)$. An increase in t raises k by dt for X and lowers it by dt for Y. The question is what happens to dq/dk as k rises.

The first-order condition is $S'(q)(q - k) + S(q) = 0$. (For future reference, note that $(q - k) = -S(q)/S'(q)$.) Total differentiation of the first-order condition yields:

$$\Big(S''(q)(q - k) + 2S'(q)\Big)\frac{dq}{dk} - S'(q) \equiv 0\,.$$

Observe $dq/dk > 0$. Total differentiating this last expression yields:

$$\left[\Big(S'''(q)(q - k) + 3S''(q)\Big)\frac{dq}{dk} - 2S''(q)\right]\frac{dq}{dk} + \frac{d^2q}{dk^2}\Big(S''(q)(q - k) + 2S'(q)\Big) \equiv 0\,. \qquad (15)$$

Multiplying both sides of (15) by $\Big(S''(q)(q - k) + 2S'(q)\Big)$ and making the substitution $q - k = -S/S'$, one obtains:

$$\left[-S'''(q)S(q) - S''(q)S'(q) + \frac{2\big(S''(q)\big)^2 S(q)}{S'(q)}\right]\frac{dq}{dk} + \frac{d^2q}{dk^2}\Big(S''(q)(q - k) + 2S'(q)\Big)^2 \equiv 0\,. \quad (16)$$

The assumption $S''' \geq 0$ is sufficient for the term in square brackets in (16) to be negative, which means that $d^2q/dk^2 > 0$.

Consider now minimizing $p + r$ with respect to t. At $t = (m_Y - m_X)/2$, $d(p + r)/dt = 0$. If t is greater, then $dp/dt > -dr/dt$, so $d(p + r)/dt > 0$. If t is less, then $dp/dt < -dr/dt$, so

$d(p+r)/dt < 0$. Hence, $t = (m_Y - m_X)/2$ minimizes $p + r$. ■

Proof of Proposition 7: Carrier Z's profits are equal to

$$S_A(p_X + p_Y)S_B(r_X + r_Y)(p_Z + r_Z - m_Z + \gamma(Z)t) \tag{17}$$

where $Z = X, Y$ and $\gamma(Z)$ is equal to -1 if $Z = X$ and 1 if $Z = Y$. By Assumptions 2 and 3, the first-order conditions for p_Z and r_Z,

$$S_A'(p_X + p_Y)\Big(p_Z + r_Z - m_Z + \gamma(Z)t\Big) + S_A(p_X + p_Y) = 0 \tag{18}$$

and

$$S_B'(r_X + r_Y)\Big(p_Z + r_Z - m_Z + \gamma(Z)t\Big) + S_B(r_X + r_Y) = 0\,, \tag{19}$$

are necessary and sufficient.

It follows immediately that $p_X + r_X - m_X - t = p_Y + r_Y - m_Y + t$ in equilibrium.

To see that the equilibrium message exchange level is independent of t, let $(p_X^0, p_Y^0, r_X^0, r_Y^0)$ denote a solution to the first-order conditions when $t = 0$. Then straightforward calculations show that, given any interconnection charge t^*, the prices $(p_X^0 + t^*, p_Y^0 - t^*, r_X^0, r_Y^0)$ satisfy the corresponding first-order conditions. The resulting level of message exchange is clearly identical to that under the original prices.

Lastly, to see that the equilibrium is unique, suppose counterfactually that there are two equilibrium message exchange levels for $t = 0$. Then it must be possible to label the price such that $p_X^1 + p_Y^1 > p_X^0 + p_Y^0$ or $r_X^1 + r_Y^1 > r_X^0 + r_Y^0$ or both. Suppose $p_X^1 + p_Y^1 > p_X^0 + p_Y^0$. (A similar logic applies if we consider receive prices.) Given that $S_A' < 0$ and $S_A'' \le 0$, the first-order conditions for p_Z imply that $p_Z^1 + r_Z^1 < p_Z^0 + r_Z^0$ for both carriers. Given that $S_B' < 0$ and $S_B'' \le 0$, the first-order conditions for r_Z then imply that $r_X^1 + r_Y^1 > r_X^0 + r_Y^0$. But it impossible to have $p_X^1 + p_Y^1 > p_X^0 + p_Y^0$, $r_X^1 + r_Y^1 > r_X^0 + r_Y^0$, and $p_Z^1 + r_Z^1 < p_Z^0 + r_Z^0$ for both carriers. Therefore, the solution to the first-order conditions must be unique.

■

References

Armstrong, Mark, "Network Interconnection in Telecommunications," *Economic Journal*, May 1998, *108* (448), 545–564.

Atkinson, Jay M. and Christopher C. Barnekov, "A Competitively Neutral Approach to Network Interconnection," December 2000. OPP Working Paper Series #34, Federal Communications Commission.

Baxter, William F., "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law & Economics*, October 1983, *26* (3), 541–588.

Bertrand, Joseph Louis François, "Review of *Theorie mathematique de la richesse sociale* par Léon Walras and *Recherches sur les principes mathématiques de la théorie des richesses* par Augustin Cournot," *Journal des Savants*, September 1883, pp. 499–508.

Cournot, Augustin A., *Researches into the Mathematical Principles of the Theory of Wealth*, New York: Kelly, 1838. English translation of French original.

DeGraba, Patrick, "Bill and Keep at the Central Office as the Efficient Interconnection Regime," December 2000. OPP Working Paper Series #33, Federal Communications Commission.

_ , "Bill and Keep Interconnection for Competing Networks," 2000. Working Paper, Federal Communications Commission.

_ , "Efficient Cost Sharing and Interconnection Rates," 2001. Working Paper, Charles River Associates.

Federal Communications Commission, *In the Matter of Developing a Unified Intercarrier Compensation Regime* 2001. Notice of Proposed Rulemaking, CC Docket No. 01–92 (rel. April 27, 2001).

Hahn, Jong-Hee, "Nonlinear Pricing of a Telecommunications Service with Call and Network Externalities," 2000. Department of Economics Working Paper, Keele University, Keele, Staffordshire, England.

Hermalin, Benjamin E. and Michael L. Katz, "On the Electronic Exchange of Messages: Sophisticated Users, Technological Progress, and Welfare," 2001. Working Paper, University of California, Berkeley.

_ **and** _ , "Retail Telecommunications Pricing in the Presence of External Effects," in Gary Madden, ed., *International Handbook on Emerging Telecommunications Networks*, Cheltenham, UK: Edward Elgar Publishing, 2001. Forthcoming.

_ **and** _ , "Sender or Receiver: Who Should Pay to Exchange an Electronic Message?," 2001. Working Paper, University of California, Berkeley.

Kaplan, Todd R. and David Wettstein, "The Possibility of Mixed-Strategy Equilibria with Constant-Returns-to-Scale Technology under Bertrand Competition," *Spanish Economic Review*, 2000, *2* (1), 65–71.

Katz, Michael L. and Carl Shapiro, "On the Licensing of Innovations," *RAND Journal of Economics*, Winter 1985, *16* (4), 504–520.

Laffont, Jean-Jacques and Jean Tirole, *Competition in Telecommunications*, Cambridge, MA: MIT Press, 2000.

Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole, "Network Competition I: Overview and Nondiscriminatory Pricing," *RAND Journal of Economics*, Spring 1998, *29* (1), 1–37.

_ **, Scott Marcus, Patrick Rey, and Jean Tirole**, "Internet Interconnection and the Off-Net-Cost Pricing Principle," 2001. Working Paper, IDEI, Toulouse, France.

Littlechild, Stephen C., "Two-part Tariffs and Consumption Externalities," *The Bell Journal of Economics*, Autumn 1975, *6* (2), 661–670.

Ramsey, Frank P., "A Contribution to the Theory of Taxation," *Economic Journal*, 1927, *37*, 47–61.

Rohlfs, Jeffrey, "A Theory of Interdependent Demand for a Communications Service," *The Bell Journal of Economics and Management Science*, Spring 1974, *5* (1), 16–37.

Schmalensee, Richard, "Payment Systems and Interchange Fees," 2001. National Bureau of Economic Research, Working Paper 8256.

Squire, L., "Some Aspects of Optimal Pricing for Telecommunications," *The Bell Journal of Economics and Management Science*, Autumn 1973, *4* (2), 515–525.

Srinagesh, Padmanabhan and Jiong Gong, "A Model of Telecommunications Demand with Call Externality," 1996. Unpublished manuscript, Bellcore.

Taylor, Lester D., *Telecommunications Demand in Theory and Practice*, Dordrecht: Kluwer Academic Publishers, 1994.

Willig, Robert D., "The Theory of Network Access Pricing," in Harry M. Trebing, ed., *Issues in Public Utility Regulation*, East Lansing: Michigan State University, 1979.