Efficient Interconnection Charges and Capacity-Based Pricing[1]

D. Mark Kennet[2] & Eric K. Ralph[3]

Telecommunications Policy Research Conference, George Mason University, VA, USA;
3 September 2006; Updated 8 February 2007

Abstract

Capacity-based interconnection (CBI) prices vary exactly with the costs a network provider incurs when supplying an interconnecting party. That is, they equal incremental costs, rather than being averaged over any output measure. We argue such prices (1) are as practicable and more efficient than per minute rates based on long run incremental cost, (2) are more efficient than bill and keep, and (3) with mark-ups for cost recovery, are a practical and relatively efficient means of pricing wholesale interconnection services, being well-suited to both circuit and packet-based networks.

1 Introduction

In terms of cost causation, a demand for interconnection amounts to a demand for capacity. Capacity-based interconnection (CBI) sets prices equal to the incremental cost of the capacity required to carry the interconnecting party's traffic.[4] Such prices are not expressed as per call or per time unit rates, but exactly as the incremental cost they reflect, that is, as a fixed fee for a specified amount of bandwidth. Thus, CBI charges the interconnecting party, on a take it or leave it basis, the full cost of the increment of capacity necessary to meet the interconnecting party's demand. For higher or lower demand, the quoted price is not scaled linearly with any output measure, but rather, is always based on the incremental cost of the desired capacity.

Proposing CBI is not new,[5] but such prices are rarely considered in practice, let alone implemented. Yet, in our view CBI prices are (1) as practicable and more efficient than per time rates based on long run incremental cost, (2) more efficient than bill and keep (assuming bill and keep is not more efficient merely because it avoids price setting and billing costs); and (3) with appropriate mark-ups for cost recovery, a practical and relatively efficient means of pricing wholesale interconnection services, including on converging networks.

---

[1] Thanks go to Bill Sharkey, Bridger Mitchell, Ingo Vogelsang and an anonymous referee for their insights, and to Henry Ergas and Scott Marcus and Ingo Vogelsang respectively for the examples from Venezuela and Germany. The authors blame any errors on each other.

[2] Telecommunications Management Group, Inc., 1600 Wilson Blvd., Suite 710, Arlington, VA 22209; www.MarkKennet.com.

[3] EKonomics LLC, 2323 Audubon St, New Orleans, LA 70125-4117; www.EKonomicsLLC.com.

[4] The paper takes it as given that a case has been made for interconnection regulation.

[5] See Vogelang, 2006 and the discussion of CBI in practice in Section 5 below; see Yoon, 2006, for a model with CBI in the context of two-sided networks; in the context of broadband see Mitchell and Vogelsang, 1991, 259-262, and Levin, 1988.

The paper is organized as follows. Section 2 outlines network costs and considers CBI prices. Section 3 considers first and second best *retail* prices. Focusing on retail markets is helpful because demand for interconnection (wholesale) services is derived from demand for retail services. As a consequence, understanding what is efficient at the retail level, as well as the problem of monopoly, which wholesale price regulation seeks to address, provides a framework for analyzing wholesale prices. Section 4 discusses first and second best wholesale interconnection prices and compares CBI with other forms of interconnection. Section 5 provides examples of CBI pricing as applied in practice.

2 Capacity Based Interconnection (CBI) Pricing

CBI requires that interconnection prices reflect their incremental costs. Interconnection minutes only materially affect costs in the way they affect peak demand. Thus CBI interconnection rates apply to peak minutes. CBI pricing is quite different to standard regulatory prices. These are based on usage (that is, per call or per time unit prices) and typically treat all priced units identically. Consequently, the resulting revenues do not bear a direct or even good relation to costs, which are determined by peak usage.

The increased availability of bottom-up engineering-economic cost models of telephone networks has made network engineering more accessible to economists (see, for example, Gasmi, Kennet, Laffont, and Sharkey, 2002). Such models reveal a great deal about what drives network costs, and are particularly important for the making of regulatory policy.

From a purely economic standpoint, network costs are typically divided into three broad groups:

1. costs that vary with the volume of traffic that can be carried during a period of maximum demand. Examples include much of the cost of switching, and trunking, notably with respect to the bandwidth of transmission links, whether over a line or carried on electromagnetic spectrum. In contrast, the underlying infrastructure used for trunking, such as conduit and towers, is relatively insensitive to the volume of peak hour traffic. The costs of traffic sensitive elements is generally measured in units of capacity, for example, E1s, T1s, erlangs, ccs, or MHz.

2. costs that do not vary with traffic volumes, but do vary with customers. On fixed line networks, the customer access loop is the most important element of this cost category; on mobile networks, handset and line card costs dominate this category.

3. costs that by and large do not vary with traffic volumes or customers. Examples include the infrastructure, such as conduit, used to carry wire between network nodes, and the buildings used to house switches, microwave equipment, cellular antennas, and the like.

Distance weakly impacts on each of these three costs groups, and for that reason is not considered in this paper.

The vast bulk of costs fall into the second and third categories. For example, the first author's cost model for wireless networks, when applied in a developing country context, shows that just under 97 percent of total costs do *not* vary with peak-hour traffic amount.[6] Similarly, in the case of fixed networks, the costs of outside plant pairs and line cards represent a substantial percentage of total costs and common to many services (Gabel and Kennet, 1991).

Figure 1 illustrates these cost categories for the case of a fixed line network.

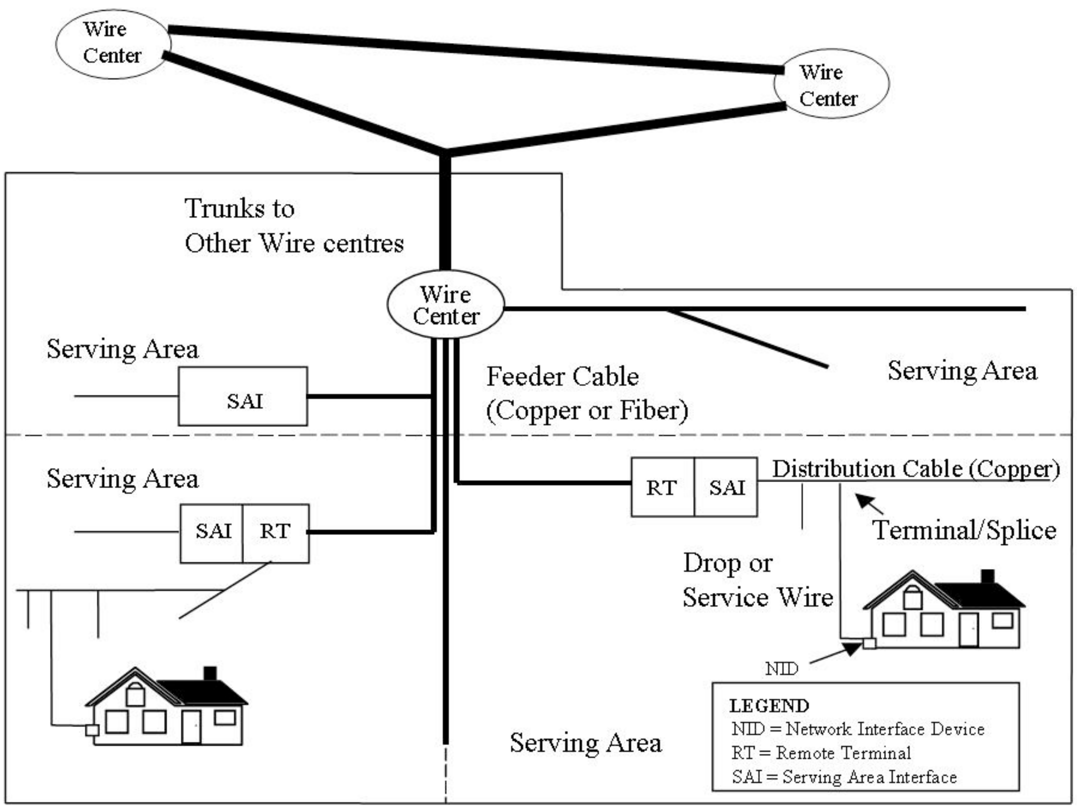

Figure 1—Wireline Network

Distribution cable, drop wires, and line cards at the wire center (not shown) can be directly attributed to individual users, but do not materially vary with telephony traffic (though broadband demand may require line upgrades). The switch cores, feeder cable, and trunks between switches are built to meet capacity requirements, that is, are traffic sensitive. Most of the remaining network elements – infrastructure – do not vary with voice or broadband traffic or customers.

---

[6] These results are not necessarily typical, but represent data from the developing world using reference element cost values. The model found that about 3.5% of total network costs are represented by incremental spectrum costs. These are the only costs that vary with peak busy-hour traffic. The remaining costs are present even when there is *de minimus* traffic.

Traffic sensitive costs vary with peak loads, that is, efficient carriage requires increased capacity when peak loads rise. The incremental cost of interconnection traffic is the cost a network owner incurs to ensure that traffic can be efficiently carried on its network, that is, the difference in costs between a network designed to meet peak demand for all traffic, and a network designed to meet peak demand for all traffic less the interconnecting traffic. Such a cost calculation is difficult because the peak depends on all traffic in the network, not merely when the interconnecting carrier's traffic peaks. Consequently, cost estimation will require some approximation that may be refined through periodic review.

The incremental cost of interconnecting traffic does not depend on whether the entrant created the interconnecting traffic or has merely claimed traffic from the network provider (so that the network carries exactly the same traffic the network provider carried pre-entry). The incremental cost is merely the cost of the traffic increment regardless of its origin, that is, the difference between the cost of an efficient network with and without the interconnecting traffic.

An example may help make this clearer. If, without the interconnecting traffic, a network has a peak in traffic at 6 pm of 100 MHz, and, with the interconnecting traffic, a peak in traffic of 110 MHz (perhaps at a time other than 6 pm), then a CBI price would charge the interconnecting network the difference between the cost of building a network designed to carry 110 MHz, and one designed to carry 100 MHz of traffic. The resulting price would be the total cost of a given expansion, which might be expressed in terms of dollars per the expansion in peak-hour capacity expansion per the expected lifespan of equipment installed (say equal to the lifespan of the longest lasting piece of equipment necessary for such an expansion). Thus, if it costs $100 to efficiently add 10 MHz to peak hour capacity, and to maintain the capacity addition that expenditure would not have to be re-incurred at the end of one year, then the incremental capacity cost could be expressed as $100/10 MHz/1 year. While this price could be represented in terms of smaller units, for example, per MHz per year, such a price is not the incremental cost of one MHz per year. This is because the incremental cost of adding one MHz for a year would typically not be linearly related to the incremental cost of adding 10 MHz for a year. For the same reason, a cost that involves at least one asset with a, say, 12 month lifespan, cannot be converted into a monthly rate by dividing by the annual price by 12.

2.1.1 From costs to prices and market supply of derivative services

One of the greatest, but often unrecognized, difficulties faced by regulators in pricing interconnection is translating cost estimates into prices applicable to periods of time that are typically much shorter than the lifespan of the relevant assets. For example, assume a zero time preference for money (to avoid having to discount) and consider a transit line that can carry traffic for ten years with a (long run incremental) cost of $100. Further assume that a carrier wishes to use the entire capacity of the line for the line's first year. Should the access price for that year be set to $10 (= $100/10)? Given uncertainty about the asset's economic value (use) over time, the answer is no. The transit line owner may have invested in the line estimating that there was a fifty percent chance that it would earn revenues of $150 over 10 years, and a fifty percent chance that it would be replaced by a cheaper technology, and would only earn $50. If access seekers could purchase at

$10 in any given year, then they in effect have an option unavailable to the investor. They can access the asset at its average valuation without taking the risk that it will be sharply devalued. If that event occurs, they can switch to the new cheaper technology, without paying the (efficient) premium that makes the investment economic. The obverse point is that a $10 per year price guarantees that the transmission provider cannot, in an expected value sense, recover its costs (even though the investment is efficient). Its expected return is not $100, but $75, being the average of the regulated return under the bad ($50) and the good scenarios ($100).

Unfortunately, it is difficult to determine short run prices that would grant facility providers with an appropriate return, and such price estimation requires substantial discretion. As a result, both regulatory error and bias will be large. To avoid these difficulties, we propose that the regulator imposes CBI prices with timeframes that reflect the expected life of the most long-lived asset. Access seekers would then have to commit to those prices on a "use-it-or-lose-it" basis, purchasing a kind of indefatigable right of use, good for the life of the capacity (call this a long run capacity right). This has the effect of allowing instant entry, rather than having to spend years rolling out network infrastructure, but still requires entrants to sink investments that parallel the incremental cost of their demand.

Despite the sunk nature of CBI prices for long run capacity rights, it is our view, that once long run capacity rights can be purchased at efficient prices, the market would also supply, if it were efficient, capacity for considerably shorter time frames (on trading in capacity rights, see Vogelsang, 2002). This would occur if there were enough independent purchasers of long run capacity rights to ensure at least some moderate degree of competition. That in turn could only happen if purchasers did not expect post-entry prices for capacity to fall to levels that would make sunk cost recovery impossible (otherwise insufficient entry would occur; Sutton, 1991, notably chapter 1). In our view, product differentiation in telecommunications is the norm, and would allow firms engaged in at least moderately effective competition to recover their sunk costs. For example, long distance carriers have long differentiated their products through branding, customer support, bundling buckets of minutes, capped call charges, classifying times-of-day, and bundling minutes to different locations. Such differentiation provided suppliers with the ability mark-up prices over marginal costs, including by use of price discrimination. This in turn ensured long distance carriers could recover fixed and often sunk costs (for example, associated with transmission lines and customer acquisition).[7] Yet, extranormal profits do not seem the norm in long distance supply. Similarly, mobile and entrants on unbundled local loop face substantial sunk costs, yet manage to recover these in relatively competitive environments without obvious extranormal profits by differentiating their services and gaining contributions by price discrimination.

[7] Over time, long distance competition might result in the commoditization of long distance minutes. Indeed, this may have occurred in countries like the US. If so, the process took well over a decade, and has led to new forms of differentiation, for example, through bundling long distance with access and other services.

Purchasers of long run capacity rights might include investment banks and other market makers who would repackage these into shorter-term packages. This might be profitable, for example, if some smaller carriers find it difficult to obtain funding to purchase long run capacity rights, and the investment bank can spread that risk over a number of small carriers.

A similar process would occur for traffic on new generation networks. There, a very great variety of services with different speeds, latency and jitter are demanded. This in turn results in wholesale demand for interconnection with many specific characteristics. However, CBI prices, as set by regulators, should only cover long run capacity rights, leaving the market to provide more complex packages. Regulators are in no position to determine the complex commercial requirements that are going to emerge on new generation networks. However, so long as basic capacity rights are available at efficient prices, the market can provide more sophisticated forms of interconnection.

3 First and second best retail prices

The preceding section considers capacity based or incremental cost prices *per se*. However, it does not demonstrate what efficient prices would look like, especially once one allows for practicalities such as cost recovery. This section briefly summarizes the relevant economic theory of efficient retail prices, as applied to telecommunications. Its purpose is to provide a framework, to be translated to wholesale prices in Section 4 below, for understanding efficient prices. An initial focus on retail prices is helpful in two respects. First, it allows questions of industry structure, notably of competitive supply (which must necessarily be addressed in the context of wholesale pricing), to largely be set aside. Second, it allows treatment of telecommunications as a standard commodity, albeit with some important externalities. In contrast, when the focus shifts to competing suppliers, it is more helpful to consider telecommunications markets as being two-sided (defined in Section 4.2 below). This change in focus, while complicating the analysis, does not change the conclusions drawn in this section.

The section starts by considering retail prices in telecommunications that achieve full economic efficiency (that is, first best prices). Our purpose is only to identify *characteristics* of retail prices that are likely to improve economic efficiency. We take such a modest approach, of only focusing on the broad characteristics of efficient prices, because it is unlikely to be within our or a regulator's power to identify, let alone implement, first best prices. Consequently, failure to gain first best prices (if such a claim could be verified, and it probably could not) is not evidence of market (Demsetz, 1969) or regulatory failure.

Retail pricing designed to ensure efficient cost recovery while minimizing inefficient bypass is then considered. Such prices are referred to as second best, because they are efficient subject to constraints (in this case, cost recovery and minimization of efficient bypass). Both first and second best prices are covered in Subsection 3.1; the views expressed there briefly are explained in detail in Section 3 of Kennet and Ralph, 2007.

The section concludes (in Subsection 3.2) with a reminder that competition commonly solves the complex pricing problems just discussed, even though regulation cannot. As a result, if wholesale regulation can allow effective retail competition, then relative to regulation, efficient prices may emerge in retail markets.

3.1 First and second best retail prices

The well-known textbook requirement for first best efficient prices is that they reflect (at least at the margin) marginal social costs, being marginal private costs plus any marginal external costs or benefits. In telecommunications, there are two important externalities than can lead to relevant marginal social costs: the subscriber and calling externality. The subscriber externality arises because existing telephone subscribers on average benefit when another person joins the network. The calling externality occurs when calls are only priced to one party to the call but both parties benefit from the call. For example, when the calling party only pays for a call (as is conventional for most fixed line calls), then the receiving party may benefit from the call without making any payment for it (the reverse may hold if the receiving party only pays).[8]

It is our view that, in practice, the subscriber and calling externalities need not be directly accounted for in setting first best interconnection prices. First, a wide range of market conduct, both commercial and due to private actions by consumers, internalizes the subscriber externality, and in the case of the calling externality, likely largely eliminates it.[9] Second, where the subscriber externality is material, it can generally be dealt with most efficiently by directly subsidizing customers (and only those customers) who are unwilling to pay the private cost of connection, but who are efficient to connect. An exception to this may be the case of networks with substantial room for growth (whether a traditional network in a poor country where penetration is still low, or a new, say, 3G network), as prices set to marginal private cost may hinder network roll-out. In these circumstances, it may be difficult for a regulator to identify who should be efficiently subsidized. However, the market still might be expected to do better. Evidence from mobile and other telecommunications supply with some degree of competition suggests that market forces can provide effective means of subsidizing network rollout. A particularly striking example of such market behavior is provided Section 4.2.1. Since firms' capacity to subsidize subscription decisions is intimately linked with interconnection policies, we postpone this discussion to Section 4, on wholesale regulation.

First best efficiency cannot be achieved unless total cost, that is, the incremental cost of total demand, is funded. If total costs are not covered, then future investment in regulated industries *economy-wide* will be inefficiently low (since failure to allow appropriate cost

[8] Calling party pays (CPP) and receiving party pays (RPP) are symmetric. Neither brings benefits the other does not, unless there is a specific demand difference on one side of the market (see Section 4.2 below on two-sided markets). The chief argument in favour of CPP only (rather than RPP only) is that on average the calling party values the call more than the receiving party. CPP and RPP are not mutually exclusive, and if both CPP and RPP are jointly used, call providers have two rather than one pricing instrument. This, in the presence of competition, would likely improve economic efficiency.

[9] In any case, regulators do not have the information required to directly internalize the calling externality.

recovery in one regulated industry signals such expropriation may occur in other regulated industries). In telecommunications, prices set equal to marginal or incremental cost will almost never recover total costs. This is because, in general, many costs are incurred prior to the priced increment (referred to as inframarginal costs) that are high relative to that increment's cost. When that is true, prices set to marginal or incremental cost neither signal nor recover the cost of total supply. This implies that even if the network and calling externalities do not need to be reflected in first best prices, marginal cost pricing cannot achieve first best efficiency.

It is true that first best prices need not achieve cost recovery, since efficiently funded government subsidies may be used instead. However, fully efficient taxes are not plausible. Consequently, in practice cost recovery requires that some prices exceed marginal cost. This leads to second best pricing—the attempt to maximize economic efficiency subject to constraints that cannot be satisfied on first best terms. The second best approach to cost recovery is to apply generalized (that is, allowing for nonlinearities) Boiteux-Ramsey prices. Unfortunately, identifying such prices is probably beyond the capacity of any regulator.

Despite this, the underlying principles of Boiteux-Ramsey prices can be used to inform regulatory pricing decisions. The Boiteux-Ramsey idea is to mark prices up above their marginal cost so that total costs are recovered in a way that minimizes the efficiency losses of those mark-ups. This essentially involves setting the highest mark-ups on products that (1) are the least responsive to demand, (2) cause the least response in demand for other products, and (3) are less likely to result in inefficient entry. Thus distortions due to the mark-ups are kept as small as possible given cost recovery must be effected.

All this implies a rather demanding program. Allowing for regulatory ignorance, the only realistic expectations one can have for Boiteux-Ramsey pricing is that regulatory decisions be informed by the basic principles that underlie such pricing, notably the three points just outlined.

3.2 Optimal retail pricing: a coda

The preceding discussion might be summarized as saying that it is too hard to regulate either first or second best retail prices. In contrast, retail pricing in effectively competitive markets can be remarkably efficient. This suggests that if regulation can enable effective retail competition, without excessively distorting input prices, then perhaps the difficulties a regulator faces are less than one might initially imagine.

Competitive markets routinely recover declining costs, often through complex discriminatory prices. As an example, think about restaurants and imagine you had no experience of competitive restaurant supply beyond some price and cost data. Would you consider alcohol and desserts to be overpriced? Would it surprise you that the relative prices of main courses at best only loosely reflect their relative marginal cost? Or that in many places bread or chips and salsa are free? Is it right for the bar to have a happy hour when prices are lower than other times? Does it make sense for bartenders to routinely receive tips of twenty percent or more when dealing directly with customers, while otherwise receiving, along with waiters, something less than 20 percent? If you thought these things were efficient, would you conclude the restaurant market in a busy section of Paris was inefficient because it had a very different set of norms?

Whatever your answers to these questions, it is unlikely that you, as a regulator, would do better by setting prices unconstrained by market forces (as anyone who experienced a Soviet block restaurant would attest). Rather, we suggest that whatever failings of the restaurant markets in New Orleans or Paris, they are superior, in a large part due to being highly competitive, to whatever would emerge if a regulator determined prices.

The relative effectiveness of retail competition in recovering costs motivates the next section of this report, which seeks to identify wholesale price regulation that will allow effective retail competition.

4 Wholesale regulation

The Boiteux-Ramsay problem, when expanded to include wholesale prices, is considerably more complex than setting only Boiteux-Ramsey retail prices (for an illustration see Laffont and Tirole, 1993, pp. 255 ff). This is all the more so when there is market power downstream, since upstream prices have to be set allowing for downstream mark-ups (for a discussion see Laffont and Tirole, 2000, pp. 124-127). Since we have previously argued that setting retail Boiteux-Ramsay prices is unrealistic, this implies that setting wholesale, including interconnection, prices that are consistent with retail Boiteux-Ramsay prices is even less achievable.

However, our pessimism about the possibility of setting first or second best wholesale prices does not imply we should forego regulating wholesale prices. Indeed, quite the opposite. To concatenate quotes, "The most important innovation (in recent regulatory policy) has been the realization that there is no compelling reason why a monopolist should have the exclusive right to use its distribution network" (Foster, 1992, p. 167). Rather, obligations to provide wholesale access can "come as close as conceivable to making the provision of telephone services at retail perfectly contestable and therefore regulation of the retail rates simply unnecessary" (Kahn, 1998). The expected result is:

(1) a substantial elimination of retail market power, simplifying the regulator's wholesale pricing problem (inefficient retail mark-ups do not have to be offset), and

(2) retail prices that are far more efficient than any regulator could ever hope to achieve. As the then chief economist at the FCC put it, "[s]moothly functioning wholesale regulation (sharing the incumbent's network…) permits and indeed almost demands retail deregulation" (Farrell, 1997).

In addition, a focus on wholesale prices has at least two more informational advantages over regulating retail prices:

(1) Wholesale regulation seeks to curb market power closer to its source, that is, the access network, which ought to make it easier to understand how efficient prices should be set. In particular, wholesale prices should not be overly influenced by retail pricing schemes, but instead should reflect wholesale costs. Thus, just as facility-based firms face capacity-based costs, even if they set retail prices according to some other metric, so to should wholesale customers face wholesale prices that reflect capacity-based costs.

(2) The range of regulated services and wholesale customers are both considerably smaller than in retail markets. For example, wholesale customers likely number in the thousands, if not less, while retail customers typical come in millions. Moreover, it would not be uncommon to find that (maybe less than) tens of customers purchase the bulk of interconnection services. As a consequence, prices that cover the marginal private cost of individual wholesalers are determined at a much higher level of traffic aggregation than prices that cover the marginal private cost of individual consumers. This has two effects on the question of ensuring inframarginal demand is efficient. First, while it remains possible that some aggregation of wholesalers' demands would not cover the marginal private cost of supplying that aggregation, the degree of under-pricing is less pronounced as compared with individual consumers. Second, a regulator wishing to set efficient inframarginal prices faces a considerably smaller number of aggregation combinations than it would in the case of retail prices.

The simplifications do not, of course, solve the wholesale pricing problem. Effective regulation of wholesale prices still requires a good understanding of (1) efficient retail prices (hence the discussion of Section 3), and (2) of how wholesale prices impact on the retail layer—the broad topic of the rest of this paper—so the regulator can meaningfully seek efficient retail outcomes.

In what follows, one-way and two-way interconnection are defined (respectively in Subsections 4.1 and 4.2). Subsection 4.1 shows that, in the simple case of a single mature network, CBI pricing is likely to be a good approximation for first best one-way interconnection pricing. Cost-recovering second best one-way interconnection prices are then considered.

When there is more than one access network, then two-way interconnection becomes necessary and this increases the complexity of first and second best interconnection prices. These issues are discussed in Subsection 4.2. A discussion of the implementation of CBI concludes the section (Subsection 4.3).

4.1 One-way interconnection

The supply of origination and termination by an owner of an access network to a carrier that does not own an access network is called one-way interconnection. The term one-way interconnection is used because one side of the market only supplies access. The canonical example is the supply of interconnection to long distance only carriers.

In considering setting one-way interconnection prices, it is initially helpful to assume a single mature access network, and that targeted retail subsidies designed to internalize the subscriber externality are in place. An example might be the supply of interconnection services to domestic long distance carriers by a national monopolist prior to the development of mobile and other alternative network technologies. In this circumstance, if efficient taxation funds any (efficient) operating deficit of access network supplier, then first best interconnection prices signal private costs, thereby allowing effective downstream competition. Put another way, in the absence of any externalities and the need to recover total costs, one-way interconnection is just like any other input. As discussed in Section 2, CBI prices are intended to exactly reflect marginal private costs, that is, the marginal cost of the capacity that can be attributed to a wholesale customer due to the bandwidth that customer demands. As a result, in the simple case of one-way interconnection on a mature network, CBI prices likely approximate first best optimal prices.

4.1.1 One-way interconnection and bill and keep

In the past decade, an increasing number of commentators have recommended that network providers do not charge each other for use of their networks—that is, interconnection charges be set to zero. Network providers are to instead recover costs from their retail customers only, a form of interconnection referred to as "bill and keep" or "sender keeps all".

Bill and keep is not suited to one-way interconnection, providing no obvious mechanism for pricing the service. Without an alternative mechanism, long distance carriers would simply have free access to origination and termination services. Bill and keep also fails to provide a means of pricing transit services, since, in many cases, the transit provider would have no relationship to the retail customer on either end of the call transiting over its network (we are grateful to Bridger Mitchell for reminding us of this). In many instances, however, transit is competitively supplied.

Free use of other carriers' networks would likely be highly inefficient. For example, long distance carriers would not incur the full incremental costs of long distance traffic, and this would bias market provision toward unbundled long distance provision. Moreover, since long distance provision is typically highly competitive, there would be inefficient overuse of long distance. In addition, bill and keep reduces the range of prices that can be used to fund inframarginal cost recovery, so is additionally inefficient on second best grounds.

One possible response would be to move away from one-way interconnection (transit aside). While that might end standalone long distance service, in some markets, most notably the US, this is already occurring despite extremely low interconnection prices. Moreover, even though long distance competition has played an extremely important role in lowering calling prices, standalone long distance service was probably an artificial regulatory creation. There are substantial economies of scope between access and calling services, both in network costs and customer management, and some, if lesser, economies in consumption, for example, the benefit of dealing with a single supplier. Given these, separate supply of line rental and long distance may be inefficient.

Alternatively, if one wishes to maintain competition from long distance only suppliers, as well as to price transit, then a one-way pricing mechanism is required. This would substantially reduce a key benefit of bill and keep, which is exactly that interconnection is not priced. More pointedly, it is argued in Section 4.2.3 below that CBI prices are superior to bill and keep in the context of two-way interconnection, but if CBI prices are adopted for two-way interconnection then it is efficient both from an administrative and incentive perspective to apply the same CBI prices for one-way and transit interconnection charges. This is because it is administratively easier to implement and oversee one interconnection mechanism, and because a uniform pricing approach reduces inefficient arbitrage opportunities. In short, the need for one-way access and/or transit charges may remove the rationale for bill and keep.

4.1.2 Comparing CBI prices to per call or per time unit prices

Consider per call or per time unit wholesale prices designed to recover the same costs as CBI prices (which are a reasonable approximation of how the FCC sets long distance interconnection charges, at least to the largest incumbent access providers). These never signal the actual cost of a call or a minute. Instead, they reflect the average per call or per time unit cost of the increment, and consequently exaggerate the cost of the bulk of supplied units, and understate the cost of units supplied at the busiest point in time. In contrast, CBI, by reflecting costs, places the access seeker on the same (resource cost) footing as the access supplier.

Any access price that is not equal to incremental cost, because it impacts on the vertically integrated firm's opportunity cost of retailing, distorts the pricing incentives of the vertically integrated firm, as well as of access seekers. This ultimately distorts retail prices (Armstrong, Doyle and Vickers, 1996). For example, in the special case when the vertically integrated firm and access seekers produce homogenous products, and compete on price, the vertically integrated firm, by setting an access price creates identical incentives for itself and its downstream rivals (see, for example, Hausman and Tardiff, 1995; Sappington, 2005) that are not typically consistent with first or second best efficiency. Thus the access price deeply colors the kinds of prices and forms of cost recovery that emerge in retail markets. This can be readily seen by how prices for international calls vary with termination charges. For example, calls with termination charges in the range of ten to twenty US cents per minute (typically to mobile networks) generally have very high per minute prices, and are not included in call buckets. In contrast, calls with termination charges in the range of less than two US cents per minute (typically to regulated fixed networks) generally have single digit per minute retail prices or can be purchased in pools.

Thinking in terms of per time unit prices has also probably distorted the terms of the debate. For example, in the US, concerns with the efficiency consequences of per time rates had led to calls for lower and lower rates, or just zero rates (bill and keep). But as discussed above, while bill and keep may solve some problems caused by pricing the marginal minute, it creates other problems. In contrast, CBI pricing focuses on cost causation rather than some simplifying (or perhaps simplistic) rule.

Another important advantage of CBI prices over traditional per time rates is that they signal the structure of the costs of facility-based entry. This reduces the likelihood that inefficient facility-based entry will occur, or that efficient entry will not occur. Per time unit interconnection rates, even when set to recover long run incremental costs on average, over price above average traffic volumes, for example, termination on heavy traffic call sinks, or above average traffic on a transit link. This can provide incentives for inefficient bypass that CBI avoids.[10] Similarly, in contrast to per time rates, CBI prices do not provide voice over Internet (VOIP) providers the opportunity to use 'broadband' bandwidth to provide voice services merely for arbitrage purposes (under a per time unit rate, VOIP avoids origination charges). For the same reason, CBI prices are naturally suited to dealing with different voice network and entirely new services delivered over new technologies. Consider a new generation access network that uses packet-based transmission all the way to the customer's equipment. An interconnection demand from a provider of streaming video probably could not be meaningfully priced on the same per time unit basis as a voice call, as the price scaled up by bandwidth would be prohibitive. However, prices that were not scaled would provide incentives for parties facing a relatively more expensive (in terms of bandwidth) price, to pass off their data as being the kind with the lowest price. In contrast, CBI guarantees uniformity, since the capacity units do not depend on the sort of information transmitted. Similarly, costs on mobile networks are largely driven by capacity costs. Thus CBI pricing naturally and efficiently translates into this setting.

CBI prices are not, however, superior in all respects. Per time unit rates arguably have at least two advantages over CBI prices:

1. Per unit time rates are easy to implement in the context of voice telephony—an originating or terminating minute of telephony is well defined and readily metered. In contrast, CBI prices may be harder to implement (but see Section 4.3 below). In part this is because regulators and industry participants are familiar with per time rates. This benefit will become less important with CBI price experience, and is unlikely to outweigh the long run benefits of prices that provide efficient pricing and investment signals.

2. Per unit time rates provide entrants with no exit costs. The entrant only pays for interconnection when it is required. In contrast, CBI prices for long run capacity would be sunk for the period of commitment. However, the market would likely provide spot or near spot CBI prices (that is, prices with no or limited time commitments) as discussed in Section 2.1.1 above.

In summary, when compared with per call or per time unit rates, CBI prices send efficient pricing and investment signals. The resulting short and long run efficiency gains are

---

[10] Sappington, 2005, notes that so long as Bertrand retail price adjustment can be expected to occur, access seekers will never engage in inefficient bypass.

likely to outweigh any benefits gained from familiarity with per time unit rates. Moreover, cost recovery and hence entry is unlikely to be prevented to the extent that CBI prices raise fixed, and even sunk costs of entry.

4.1.3 Second best considerations: Allowing for wholesale cost recovery

The preceding discussion suggested that one-way (typically long distance) entrants would not necessarily face difficulties in recovering their costs under CBI prices. However, CBI pricing, in general, will not allow a network provider to recover its inframarginal costs. To see this, it will be helpful to remember the example of monopolist with a mature network in developed country. As discussed in Section 3 above, first best interconnection prices would be unlikely to allow the monopolist to cover its inframarginal costs, notably the costs of corporate overheads and the access network. Instead, cost recovery requires efficient second best retail prices. However, efficient second best retail prices would not be sustainable in a market where underlying wholesale charges were not also designed to allow cost recovery. If wholesale prices purely reflected the incremental cost of capacity, then competition (which interconnection pricing is intended to generate) would likely bid away any quasi-rents necessary to cover the access provider's upstream inframarginal costs. As a result, retail competition would allow recovery of retail costs and incremental wholesale (CBI) costs only. The wholesaler could claim no contribution towards inframarginal wholesale costs.

This implies that CBI prices require mark-ups for cost recovery to achieve second best efficiency.[11] Such mark-ups will generally distort choice less than the mark-ups required over the marginal cost of an origination and termination service that does not distinguish between, for example, use of the access line and transit. Standard origination and termination charges are solely levied on a per time unit basis. Thus, to recover the services' incremental cost, the origination and termination charges must recover not only the (likely zero) marginal cost of traffic on access links, but also the generally nontrivial incremental costs of transit. To this must be added a mark-up for overall cost recovery. In contrast, the CBI approach sets separate prices for origination and termination on a given access line and transit. In the case that transit charges raise nontrivial revenues, the mark-up on access line origination and termination need not recover the incremental cost of transit, and so is more efficient (the transit cost is directly and more efficiently recovered in the CBI transit price). Finally, separate mark-ups can be applied to access line origination and termination and to transit costs, providing the regulator with more tools for minimizing welfare losses.

It is also our view that the mark-ups over CBI prices ought to be set in a manner that respects the nonlinear nature of network cost structures. This has two advantages. First, such cost-like signals are consistent with the cost structure of facilities-based supply so

[11] We assume that cost recovery must be effected through interconnection charges, rather than by a retail tax. While standard tax theory suggests a retail tax would typically be more efficient, wholesalers may be more efficient in recovering costs than a regulator (as argued in the introduction to Section 4 above). If alternative forms of access already have mark-ups for cost recovery, then efficiency requires that a mark-up be levied on all forms of access.

are less likely to lead to interconnection purchasers developing business models that are inconsistent with facilities-based models purely due to regulatory price setting. Second, nonlinear per time unit mark-ups over CBI prices that decline toward CBI prices at the margin are likely to provide more efficient signals for use and investment than linear per time unit price (which typically characterize standard origination and termination prices).

Merely arguing for nonlinear mark-ups that decline toward CBI prices, however, leaves a crucial question unanswered: that is, what contribution should particular services, say, long distance calling, make toward inframarginal wholesale cost recovery. The lessons of Boiteux-Ramsey pricing provide some guidance. A regulator could build a relatively simple *retail* model of access, local call, long distance, vertical and perhaps other services, and determine Ramsey mark-ups in that scenario. The results can then be used, assuming effective competition in the long distance market, to approximate the total cost contribution that should be obtained from the one-way interconnection service. The regulator would then estimate the implicit and actual revenues CBI prices would generate from long distance calls supplied by the incumbent to itself and to third parties, and subtract these from the estimated cost contribution. The regulator would next identify a nonlinear wholesale mark-up function that recovered the difference between the estimated cost contribution and the estimated CBI pricing revenues. Subsequently, the regulator could review actual retail cost recovery and adjust its cost contributions to reflect market outcomes.

The need to minimize inefficient bypass complicates the process. The whole mark-up determination exercise just outlined takes no account of the fact that supply, for example, in some locations or to certain types of customer, may be more likely than in others to generate bypass at the implied wholesale prices. It may therefore be more efficient to reduce the total cost contribution in those cases, and raise them in others (including beyond one-way interconnection charges).

Concluding, nonlinear mark-ups on CBI prices are in general necessary for economic efficiency. Such prices can be expected to be more efficient than per unit prices or even nonlinear prices designed to raise the same level of total revenues, but which do not have CBI costs as part of their underlying base. This is because marked-up CBI prices only incorporate mark-ups after incremental costs of all services (for example, access line origination and termination and transit) have been recovered, while even nonlinear prices for origination and termination must recover all costs on a single regulated product.

4.2 Two-way interconnection

When a call originates on one access network and terminates on another, then the networks are said to engage in two-way interconnection. The term two-way refers to the fact that the originating and the terminating network supply access to each other (in contrast, a pure long distance carrier does not supply access).

Closely related to two-way interconnection is the concept of a two-sided market (Rochet and Tirole, 2003). In a two-sided market, a platform provider connects two different types of customers with the peculiarity that each customer group cares about (roughly speaking) the size and/or degree of access it has to the other group. Thus, by joining a platform, a subscriber brings an external benefit to subscribers from the other side (since they now have access to more subscribers of the first group). Examples include magazines as a platform joining readers and advertisers (readers may not directly value advertisers, but do value the quality of content that the advertiser makes possible) and credit card providers as a platform joining purchasers and sellers.

When considering interconnection between networks, it is helpful to use a two-sided market approach. In this context, networks are competing platforms that connect callers to call recipients. Due both to the subscriber and the calling externality, subscribers on any one network benefit from the presence of subscribers on any other (for a specific example with telecommunications networks see Wright, 2002b).

The results of Section 3.1 on first best retail pricing can be readily summarized in terms of two-sided market analysis:

- First best efficient retail pricing for two-sided markets generally requires prices on both sides of the market to equal to marginal (social plus private) cost, that is, the CBI price, less the marginal benefit gained by the other side of the market (Armstrong, 2006).

- Such prices cannot recover costs when marginal costs are constant (Bolt and Tieman, 2005). Moreover, it is easy to show, for example, by adding a fixed cost, that cost recovery is even more difficult when costs fall with output.

- However, for the same reasons as given in Section 3, at least for relatively mature networks, setting retail prices to marginal private cost and applying targeted subsidies to subscribers are likely all that is required to ensure a reasonable approximation to a first best outcome.

4.2.1 First best two-way interconnection prices

On a mature network, as with one-way interconnection, targeted retail subsidies designed to internalize the subscriber externality simplify the setting of (wholesale) interconnection prices. Such subsidies, alongside first best funding of network costs not recoverable from retail prices, mean that CBI prices (which by definition reflect marginal private costs) can allow efficient competition downstream.

The regulator’s task is somewhat more complex when one network is mature, and a second is based on a new technology, or targets a group of people whose decisions to subscribe tend to be marginal. In either case, some subsidy toward subscription decisions on the second network is, in general, called for. Consider first the introduction of a new technology, say, mobile service. Fixed line customers benefit from being able to reach people who, except for their mobiles, could not be contacted. However, even wealthy consumers’ subscription decisions for a new technology are likely to be impacted by price (both because it will be initially expensive, and because potential consumers will be unsure as to the value of the new service). It is therefore plausible that discounting the marginal cost of mobile service by the benefit fixed line callers gain from the service would provide price signals that allow efficient expansion of the new service. Similarly, consider two networks, one with relatively wealthy customers who by and large would pay the marginal private cost of subscription, and another network that targets marginal customers. An example might be a network in a wealthy region or country and another network in a poorer region or country, especially if the latter also has higher marginal private costs (for example, due to lower line density). In this case, subsidizing poorer users of the second network may well be optimal.

There are, however, some practical difficulties in implementing optimal subsidies. First, the optimal subsidy must be identified. Second, once identified, the subsidy would likely need to be funded within the industry (rather than from general taxes). In that case, the customers on the mature (wealthy) network who benefit from being able to reach customers on the new network are the natural tax base (in terms of minimizing distortion of choice). Third, ideally the required subsidies should be targeted, for example, through a rebate on subscriptions to the service based on the new technology (or poor network). However, this may not always be possible if calls cross jurisdictional boundaries (for example, an international border).

Putting the third, jurisdictional problem, aside, the regulator's task in distributing subsidies is greatly eased by pricing at the wholesale layer when there is competition for customers who require subsidization (as, for example, often can be the case where competing networks roll-out service—mobile telephony provides an example). In that case, the regulator can impose a mark-up on charges to terminate calls on the networks to be subsidized. Competition by the subsidized networks for those call termination revenues, and ultimately for subscribers, will ensure that any revenues above cost that are transferred from the subsidizing networks to the subsidized networks are passed on to customers (this is sometimes referred to as the waterbed effect—see, for example, Littlechild, 2006).

Competition can also be helpful in other, sometimes surprising ways, which calls for an open mind on the part of regulators. An ultimately tragic example comes from Venezuela.[12] There, telephony suppliers, including the main mobile supplier, have traditionally targeted the wealthy, since the wealthy can afford cost-covering prices. That this was suboptimal was made obvious in practice by a new mobile network that specifically targeted poorer customers.

The new mobile service was based on the older AMPS standard, often using secondhand network equipment and handsets. Subscription prices were low, and outbound calls were sufficiently expensive that the network's predominantly poor subscribers rarely made such calls. However, inbound calls were free to the subscriber. The effect was that, even though the technology was cheap, fees from subscribers did not recover network costs.

Despite the deficit between costs and subscription fees, the AMPS network was kept profitable by substantial termination fees earned from calls originating off net. Wealthier

---

[12] We are grateful to Henry Ergas for this example.

community members made these calls from fixed lines or from modern mobile networks. Suddenly, poorer citizens could find work without, often fruitlessly, traveling long distances. Instead, for a relatively low monthly fee, work found them. Equally, from the perspective of the wealthy, gardeners, child-minders, maids, and laborers were available in a way that had been previously unimagined.

The effect was that the wealthy users of other networks were subsidizing the relatively poor users of the AMPS network to everyone's benefit. Moreover, there was little migration by the wealthy from expensive mobile networks to the relatively cheap AMPS network. Wealthy individuals preferred more expensive mobile services to the cheaper service because: (1) the AMPS service was less sophisticated, and (2) the AMPS handsets were unfashionable (an important factor among the status conscious wealthy).

While one might argue as to whether this arrangement constituted Nirvana, it seems to have offered a significant Pareto improvement over the *status quo* and illustrates not only that optimal pricing on one side of a two-sided telecommunications market might efficiently subsidize costs on the other, but that market forces may even be capable of achieving this goal. Sadly, the regulatory authorities in Venezuela considered the cheaper network's pricing plan as treating the poor unfairly, that is, as being discriminatory in a pejorative sense, and foreclosed the high termination charges that sustained the cheaper network's business plan.

4.2.2 Second best two-way interconnection prices

Cost recovery issues also apply to two-way interconnection. As before, it is unlikely that a regulator can identify full-blown second best prices, but can seek to apply Boiteux-Ramsey principles when setting prices. For example, it is likely efficient to mark (wholesale) termination charges up much more sharply than origination charges for CPP calls (and *vice versa* for RPP calls like 800 numbers). As different services, both call origination and termination should be taxed to ensure cost recovery (the tax base should be as broad as possible). Further, and for simplicity focusing on the CPP calls, while the demand elasticity for call origination is likely quite high (consider the ability of a VOIP provider or unbundled local loop provider to avoid these), demand for call termination is highly inelastic. For example, the calling party generally has a much better idea of the value of the call to themselves, than the called party has. Consequently, the calling party is more likely to be less responsive to price than the calling party. Similarly, the calling party often has very few options for quickly reaching the called party (this is especially true of calls to mobiles when the caller does not know the recipient's location, but is often also true of calls to fixed lines). Hence substitution toward another network is not likely (which is related to why mobile termination charges, even when mobile supply is highly competitive, have received so much regulatory attention in recent years). While the degree to which a network has market power over call termination is sometimes exaggerated, it is true that it is costly for retailers to set multiple prices for calls terminating on different networks, and that retail consumers often would not know what those prices are, or what network the party they wish to call uses.[13] Thus, demand for call

---

[13] In two-sided markets, if one side of the market essentially uses only one platform (referred to as single homing), and the other side many or all platforms (multi-homing), as is typical of wholesale

termination services is likely highly inelastic relative to demand for many other telecommunications services, most obviously call origination, and this calls for a higher mark-up.

4.2.3 Two-way interconnection: Comparing CBI with per call or per time unit rates and bill and keep

CBI prices are superior to standard per call or per call minute interconnection prices when there is more than one access network for the reasons given in Section 4.1.1 above. In addition, under two-way interconnection and in contrast to CBI prices, per call or per time unit interconnection prices (1) distort wholesale competition, for example, by providing incentives to inefficiently supply access to call sinks, and (2) are completely unsuited to billing demand for interconnection of new non-voice services (which do not provide calls minutes).

CBI prices are also superior to bill and keep for the reasons given in Section 4.1.2 above. Further, bill and keep can be an inefficient means of interconnecting mature networks, if they have asymmetric costs, or calling patterns. For example, bill and keep makes customers who are expensive to serve, say in less dense or more difficult to serve areas, less attractive than is efficient if they generate inbound calls. This may lead to inefficient competition for such customers, and inefficient investment in supplying such customers as compared with what the inbound calls that they generate would justify. Similarly, bill and keep leads to artificial competition for customers who make large numbers of outbound calls. Customers (under a CPP regime) who initiate a lot of outbound calls are more attractive than customers who tend to receive, but not make, calls because only the callers generate revenues. This, however, is likely to have little bearing on which type of customer is efficient to provide. For example, bill and keep network providers would prefer wealthier customers (who make more calls) to poorer customers even when allowing positive termination charges would benefit both groups (as in the example from Venezuela). See also Wright, 2002a.

However, in the context of two-way interconnection, it is interesting to note that when there is competition for interconnection services beyond the point of traffic concentration, (regulated) CBI and bill and keep (if not applied beyond the point of traffic concentration—for example, as proposed by DeGraba, 2000) are identical and consistent with first best pricing. Under CBI, when competition supplies services beyond the point of traffic concentration, the regulator only sets prices for use of the access network. Since access network costs do not vary with volume, the result is a zero interconnection charge, that is, bill and keep. However, if prices for interconnection services beyond the point of traffic concentration, such as transit services, require regulation then, CBI prices are positive and first best, while bill and keep prices are always zero and so do not reflect marginal costs.

---

telecommunications markets, then platform providers have market power over the multi-homing side (call terminators). Despite suggestions to the contrary (Littlechild 2006), market power over wholesale multi-homers does not change with whether CPP and RPP are used at the retail level. Further, regulating interconnection rates, for example, by use of CBI or bill and keep, constrains this market power.

CBI and bill and keep have another similarity. Two mature networks with symmetric structures and costs may prefer bill and keep to CBI interconnection, because traffic flows between such networks are likely to be balanced, and so the costs that each network imposes on the other would on average net out. This, however, does not provide an argument for abandoning CBI regulation in favor of bill and keep (since this would provider carriers with incentives to see out cheaper customers or customers with unbalanced traffic). Rather, CBI prices should be set subject to commercial override (as discussed in the next section). Indeed, in practice, bill and keep like agreements exist, notably for Internet interconnection, but typically are conditioned on matters such as traffic balance and network coverage. As these conditions are likely to be complex and vary with the particular carriers in question, such arrangements are best left to the interconnecting carriers.

4.3 On implementing CBI

From a practical perspective, we consider that commercially negotiated agreements should override regulated prices, even if they are capacity-based. In practice a regulator will make errors in estimating marginal costs, including in the choices made to simplify prices for similar services (for example, over the many possible different network pathways calls can transit[14]). Moreover, the regulator's capacity to estimate efficient mark-ups is limited. In contrast, effectively competitive markets are likely to set more efficient prices than regulators. Thus, it is better to specify CBI prices as a default, but to allow commercial interconnection agreements to override these (Graham and Vernon, 1991). This provides the regulated parties the possibility of choosing Pareto superior options unknown to the regulator. Of course, commercial interconnection agreements would be subject to standard competition law oversight, but in any case, collusion through interconnection agreements would generally be difficult to sustain (since the pricing dimensions of interconnection agreements can be expected to exceed the dimensions over which interconnecting carriers compete for retail customers).

A preference for commercially negotiated contracts led some countries, notably the US and Australia, to require that the relevant parties first seek commercial arrangements before regulatory intervention. When commercial negotiations fail the parties are entitled to seek arbitration from the regulator, where guidelines as to what kind of prices would be considered reasonable are known in advance. We consider this approach misguided because it provides minimal information to the regulator, and increases regulatory uncertainty. In effect, the guidelines under which commercial negotiations are to be arbitrated severely constrain the outcomes of commercial negotiations. In particular, the parties that have the most to gain under arbitration have no incentives to identify more efficient contracts. Moreover as the arbitration process can be lengthy (as it is in Australia) and its outcomes unpredictable, this greatly increases industry uncertainty.

We prefer directly mandating CBI prices, subject to commercial agreements to current practice since it provides immediate certainty, and is unlikely to constrain commercial agreements any more than the threat of arbitration with pre-announced guidelines.

---

[14] In a packet-switched network, averaging will reflect *ex ante* costs (probabilities must be assigned to the transmission paths a packet can take, since *ex ante* these cannot be known).

An arguably better option comes in the form of final offer arbitration. While this has the disadvantage of taking time and uncertain outcomes, it has the important benefit of providing the arbitrator-regulator with substantial information. Under final offer arbitration the parties each proffer a single contract to the regulator who then choose one or the other in its entirety. Final offer arbitration provides strong incentives to be reasonable, since it is the most reasonable offer that will be accepted. In contrast, ordinary arbitration processes drive the parties apart.

That being said, it is important to recognize that final offer arbitration is best suited to negotiations along a single dimension. Unfortunately, interconnection contracts are not so simple, and this is so even if only price is considered. For example, one supplier might suggest a per time unit price equal to the regulator's estimate of long run average incremental cost divided by total traffic, while the other might suggest a capacity based price equal to the same amount divided by total capacity. Moreover, it is quite probable that both sides would propose several, and not always the same, kinds of interconnection. For example, one party might wish to have different prices for calls that terminate within one exchange and calls that terminate outside of the originating exchange area, while the other might prefer to distinguish between "local" calls defined by some other regional calling area. Such complexities create incentives to differentiate bids in ways that makes it hard for the arbitrator to determine reasonableness. This in turn creates pressures to give the arbitrator greater discretionary power (since the process that led to the final offer arbitration scheme can also be used to dismantle it), thereby undermining the main advantage of the approach.

In short, final offer arbitration may be no panacea, and setting (CBI) prices that commercial negotiation may override may remain the best option.

5 CBI in practice

At this juncture, several international jurisdictions have approved some version of CBI and are at various levels of implementation. We describe briefly each country's decision and comment from the perspective of the points raised in this paper.

Spain. Spain is the European pioneer of CBI. The decision on the part of CMT in August of 2001 allowed Telefónica de España to publish a reference interconnection offer on a capacity basis (referred to as CABIS). Unfortunately, wrangling over the terms of the definition and costing of interconnection have been ongoing, and to our knowledge no CBI agreements have been signed. Telefónica has submitted to the ITU a paper describing its proposal for costing, outlining a number of perceived technical difficulties (Telefónica, 2003). In addition, the CMT has proposed that capacity requirements be estimated using a factor multiplied by actual interconnection usage. This raises a concern that if applied incorrectly, the result will be another usage-based tariff and will not lead to the benefits we describe in this paper.

United Kingdom. In the UK, call origination prices for dial-up Internet access (referred to as Flat Rate Internet Access Call Origination or FRIACO) are a form of CBI, albeit confined to a narrow and dying service (Ofcom, 2006, paragraph 1.13, has indicated the requirement to supply FRIACO will be reviewed shortly). The service, however, was less introduced because of the attractions of CBI *per se* (otherwise it might have had wider application) and more as a means of allowing dial-up access that was not subject to the standard per minute originating rates.

Portugal. In a public consultation on CBI, Anacom, Portugal's telecom regulator, essentially decided that it would permit CBI for a limited number of services, but eliminated the possibility for other services, including one ideally suited to CBI—data services, as well as long distance termination (Anacom, 2005). It would appear that Anacom does not recognize that it is simpler to regulate interconnection at the network, rather than service level. For example, this avoids the need for service definitions, and will not lead to arbitrage between services that are regulated differently or the creation of artificial services designed to avoid or obtain pricing benefits. Consequently, it is possible that Anacom's approach will yield few or even negative benefits.

Colombia. In 2000, Colombia's Comisión de Regulación de Telecomunicaciones (CRT) became a Latin American pioneer when it explicitly attempted to address the convergence issue via interconnection policy. CRT's Regimen Unificada De Interconexión (RUDI) includes a set of obligations and principles that apply to all telecommunications operators and service providers, as well as special obligations that apply to selected operators. RUDI includes the option of CBI charges calculated on the premise that the operator providing interconnection shall recover its costs of operation, maintenance of the network, plus a reasonable profit, independent of the volume of traffic. The operator that purchases capacity assumes the risks associated with traffic fluctuations. CRT also set per-minute charging alternatives, meaning that access seekers have the choice between per-minute and capacity-based rates (Tamayo, 2003). At this writing, it is unclear to what extent operators have taken advantage of the CBI alternative, or if per time unit rates are set at levels inconsistent with CBI charges. It also appears that wrangling over costing rules is holding up implementation.

Peru. Peru became the second Latin American nation to introduce the possibility of CBI in its interconnection decision of 2003. In that decision, Osiptel (Organismo Supervisor de la Inversión Privada en Telecomunicaciones) set a traditional usage-based cost-oriented rate for termination on Telefónica del Perú's landline network, but announced its openness to CBI on the request of the party requiring termination. At this writing, no firm has requested to terminate on TdP's network using CBI, nor has TdP requested such a rate on any other network. Osiptel has promised a consultation on determining a cost-based CBI tariff, but so far has not delivered.

Jordan. Similarly to the case of Peru, Jordan issued a consultation document in 2005 describing its willingness to permit CBI, but has not yet provided rules of implemention. Jordan's situation may be somewhat unique among the countries considered here in that the incumbent wireline network, Jordan Telecom, is roughly the same size as the largest wireless network, FastLink, and rivalry between the two firms is intense. If Jordan's Telecommunication Regulatory Commission implements CBI, Jordan Telecom will face an interesting decision in pursuing capacity-based arrangements. It will possibly lose revenue in terminating international traffic, but may gain it back on terminating traffic on FastLink if it can appropriately market off-net traffic. However, the Jordanian regulator's capacity to implement CBI has been hampered by the departure of its former chair, and a loss of expertise among its staff.

6 Conclusion

In spite of the promise that CBI would appear to offer, it has only been undertaken in a handful of countries, almost always in parallel with more standard per time unit pricing. Regulatory reluctance to implement CBI is probably partly explained by the novelty of the idea. Such a ground up reconsideration of interconnection is likely to require substantial resourcing and necessarily creates a degree of risk for the regulator. It is unlikely that the present test beds for CBI will realize the potential benefits of the approach for several reasons. First, the small countries that have considered CBI do not have the regulatory budgets capable of properly developing such a new approach. Second, given the parallel availability of per time unit prices, interconnection purchasers might prefer to remain with the familiar than shift to CBI. Third, the availability of per time unit prices raises the prospect that one service will (erroneously) be cheaper relative to the other. If this is the CBI price then CBI is unlikely to gain any traction. Fourth and finally, parallel pricing substantially increases the probability that one service will be priced below cost leading to inefficient oversupply.

Bibliography

Allen, James (2006). The death of 'price per minute' for fixed interconnect? Analysys (sighted 28 October 2006) http://www.analysys.com/default.asp?m=966&n=52052&k=MAR_EN_oct06_pja&Mode=article&iLeftarticle=2238

Anacom (2005). "Consultation Document: Capacity-based Interconnection Offer in Portugal." http://www.icp.pt/template12.jsp?categoryId=157130. 29 June 2005.

Armstrong, M. (2002). The Theory of Access Pricing and Interconnection. *The Handbook of Telecommunications Economics*. M.E. Cave, S.K. Majumdar and I. Vogelsang. Amsterdam, North-Holland. 1: 295-384.

Armstrong, M. (2006). "Competition in two-sided markets." *RAND Journal of Economics*.

Armstrong, M., C. Doyle, *et al*. (1996). "The access pricing problem: A synthesis." Journal of Industrial Economics 44(4): 131-50.

Bolt, W. and A.F. Tieman (2006). "Social Welfare and Cost Recovery in Two-Sided Markets." *Review of Network Economics* 5(1): 103-117.

DeGraba, Patrick (2000). "Bill and keep at the central office as the efficient interconnection regime," OPP Working Paper Series 33, December, FCC.

Demsetz, H. 1969, Perfect Competition, Regulation, and the Stock Market. *Economic Policy and the Regulation of Corporate Securitie*s. H.G. Manne. Washington, D.C., American Enterprise Institute: 1-22.

Hausman, J.A. and T.J. Tardiff (1995). "Efficient Local Exchange Competition." *The Antitrust Bulletin* (Fall): 529-556.

Farrell, J. (1997) "Prospects for deregulation in telecommunications", FCC, Washington DC, 9 May

Foster, C. (1992), *Privatization, Public Ownership and the Regulation of Natural Monopoly*, Blackwell: Oxford, U.K., pp. 167 ff.

Gabel, D. and M.D. Kennet (1994). "Economies of Scope in the Local Telephone Exchange Market." *Journal of Regulatory Economics* 64(4): 381-298.

Graham, D.A. and J.M. Vernon (1991). "A note on decentralized natural monopoly regulation." *Southern Economic Journal* 58(1): 273-5.

Kahn, A. (1998). *Letting Go: Deregulating the Process of Deregulation*, Michigan State University Institute of Public Utilities, East Lansing, MI.

Kennet, D.M. and E.K. Ralph (2007). Efficient Interconnection Charges and Capacity-Based Pricing. www.EKonomicsLLC.com/KennetRalph2007CBI.pdf

Laffont, J.-J. and J. Tirole (1993). *A theory of incentives in procurement and regulation*. Cambridge, MA, MIT Press.

Laffont, J.-J. and J. Tirole (1994). "Access pricing and competition." *European Economic Review* 38(9): 1673-1710.

Laffont, J.-J. and J. Tirole (2000). *Competition in telecommunications*. Cambridge, MA, MIT Press.

Levin, S.L. (1988). The economics and pricing of broadband. *16th Telcommunications Policy Research Conference*. October-November.

Littlechild, S.C. (2006). "Mobile Termination Charges: Calling Party Pays versus Receiving Party Pays." *Telecommunications Policy* 30(5-6): 242-277.

Mitchell, B.M. and I. Vogelsang (1991). *Telecommunications pricing: Theory and practice*. Cambridge, UK, Cambridge University Press.

Ofcom (2006). "Next Generation Networks: Developing the regulatory framework" http://www.ofcom.org.uk/consult/condocs/nxgnfc/statement/

Rochet, J.-C. and J. Tirole (2003). "Platform competition in two-sided markets." *Journal of the European Economic Association* 1(4): 990-1029.

Sappington, D.E.M. (2005). "On the Irrelevance of Input Prices for Make-or-Buy Decisions." American Economic Review 95(5): 1631-1638.

Sutton, J. (1991). Sunk costs and market structure: price competition, advertising and the evolution of concentration. Cambridge, MA, MIT Press.

Tamayo, G. (2003). "Implementing Capacity-Based Interconnection Charges to Promote Affordable Internet Access." ITU

Telefónica de España S.A. (2003). "Points to Consider when Introducing a Capacity-based Interconnection Model in a Context of Competing Networks," ITU, Telecommunication Development Bureau, Document RGQ6-1/1/14-E, 1 April 2003.

Vogelsang, I. (2002). Determining interconnection charges based on network capacity utilized. *Price regulation*. K.-H. Neuman, S. Strube Martens and U. Stumpf. Bad Honnef, WIK proceedings: 95-129.

Vogelsang, I. (2006). “Abrechnungssysteme und zusammenschaltungsregime aus okonomischer sicht.” http://www.bundesnetzagentur.de/media/archive/6202.pdf

Wright, J. (2002a). "Bill and Keep as the Efficient Interconnection Regime?" Review of Network Economics 1(1): 54-60.

Wright, J. (2002b). "Access Pricing under Competition: An Application to Cellular Networks." *Journal of Industrial Economics* 50(3): 289-315.

Yoon, K. (2006). "Interconnection Economics of All-IP Networks." *Review of Network Economics* 5(3): 351-365.