Access Pricing in Telecommunications

COMPETITION LAW AND POLICY

OECD

Access Pricing in Telecommunications

OECD

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Pursuant to Article 1 of the Convention signed in Paris on 14th December 1960, and which came into force on 30th September 1961, the Organisation for Economic Co-operation and Development (OECD) shall promote policies designed:

– to achieve the highest sustainable economic growth and employment and a rising standard of living in member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
– to contribute to sound economic expansion in member as well as non-member countries in the process of economic development; and
– to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

The original member countries of the OECD are Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The following countries became members subsequently through accession at the dates indicated hereafter: Japan (28th April 1964), Finland (28th January 1969), Australia (7th June 1971), New Zealand (29th May 1973), Mexico (18th May 1994), the Czech Republic (21st December 1995), Hungary (7th May 1996), Poland (22nd November 1996), Korea (12th December 1996) and the Slovak Republic (14th December 2000). The Commission of the European Communities takes part in the work of the OECD (Article 13 of the OECD Convention).

Publié en français sous le titre :
La tarification de l'accès dans le secteur des télécommunications

Foreword

This report on access pricing and the regulation of access services has been prepared by the OECD's Competition Committee.[*] The Competition Committee has, over the past ten years, explored various aspects of the interaction between competition and economic regulation. This work has looked both at specific sectors (such as the public utilities, banking, broadcasting and so on) and at specific issues which cut across a number of different sectors (such as issues related to vertical structural separation, or the handling of non-commercial services). This report addresses an issue which arises in virtually every public utility industry – the regulation of access to essential facilities. Access regulation issues lie squarely at the intersection of the fields of competition and regulatory policy – on the one hand access regulation draws on all aspects of the theory of regulation of natural monopolies; at the same time, as this report emphasises, both the level and structure of access prices have important implications regarding the scope for competition in the competitive parts of the industry.

This report was prepared in three parts over the period October 2001-October 2003. The first part addresses the economic theory of access pricing, drawing strongly on the theory of regulation of a natural monopoly. At the same time, OECD countries were invited to make submissions regarding how access regulation is carried out in their country, focusing on the telecommunications sector. This material is summarised in the second part of this report. Since the importance of "cost based" access prices is firmly established in conventional wisdom, the third part of this report looks at what it means for prices to be "cost based" in the presence of both fixed and common costs.

Access issues are not easy issues for regulators and competition policy-makers – they can be both difficult technically and the source of noisy disputes in practice. By bringing together theory and practice, this report seeks to help regulators and policymakers learn

---

* *The report was drafted by Darryl Biggar, consulting economist to the ACCC, previously member of the OECD Competition Division.*

Table of Contents

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Introduction

"It is hard to find a more controversial issue in industrial policy than that concerning the terms on which entrants can gain access to an incumbent firm's network."*

As a result of the waves of deregulation of the past two decades, the regulation of the terms and conditions under which competing firms have access to essential inputs provided by rivals has become the single biggest issue facing regulators of public utility industries. This issue is both theoretically complex and inherently controversial. Since the development of competition and the success of liberalisation often depend on the access terms and conditions chosen, there is also a strong public policy interest in getting these terms and conditions "right". At the same time, new entrant firms and incumbents often have a substantial financial stake in the outcome and therefore a strong interest in negotiating aggressively.

This report, prepared by the Competition Committee of the OECD, seeks to deepen our understanding of access regulation in order to better understand the trade-offs that must be made, to enhance the quality of access regulation and to enhance the scope for competition in the competitive segments of regulated utility industries. This report focuses on the telecommunications industry but almost all the results have direct application in other public utility industries.

The report consists of three parts. The first part is a survey of the theory of access pricing. This chapter identifies the basic types of access problems and some key principles to bring to the practice of access pricing. The second part makes some attempt to bring together this theory with the practice in the telecommunications industry. We see that, for various reasons, countries don't choose to differentiate access prices as much as expected under the theory of natural monopoly and, in some cases, this gives rise to competition concerns. Because so much of the discussion surrounding access pricing revolves around the idea of "cost-based" prices, the third and final part of this report looks at what it means to set "cost-based" prices. We will see there that measuring costs requires an allocation of any common costs that arise when multiple products can be produced more cheaply jointly than individually.

* Armstrong (2001), page 36.

Typically these costs are allocated according to mechanical "rules of thumb". This chapter also looks at how these costs might be allocated in a way that leads to efficient recovery of investment costs.

The key ideas of this report can be summarised as follows:

Chapter 1 – The theory of access pricing

- Access problems arise when the provision of a complete service to end-users requires the combination of two or more inputs, one of which is non-competitive (*i.e.*, a monopoly). It is possible to distinguish (at least) two different types of access problems. The first category, called "one-way" access problems, arise when providers of a competitive service need to purchase essential inputs from the provider of the non-competitive service but not *vice versa*. The second category, called "two-way" access problems, arise when each firm in the market must purchase an essential input from the other firms in the market.
- The one-way access pricing problem is very similar to the classic problem of the regulation of the prices of a natural monopoly. Many of the principles of natural monopoly regulation apply directly to the regulation of access prices. In particular, the theory of natural monopoly regulation highlights the fact that the structure of the regulated prices will depend on the need to raise revenue in excess of that raised pricing at marginal cost and the extent to which the firm or its regulator can discriminate in its prices.
- Most public utility industries are characterised by economies of scale and scope in at least some segments of the industry. In the presence of scale and scope economies setting all the prices of a firm at marginal cost would not allow the firm to earn sufficient revenue to cover its total costs. Revenue in excess of that collected by marginal cost prices can be raised in several ways: general tax revenues (or revenues drawn mostly from outside the regulated industry), taxes on the products of the regulated firm and its rivals (which, in effect, fall on both competitive and access inputs to the regulated product) and access pricing. This report assumes general revenues are not available for this purpose and, further, focuses on the use of access pricing. Where general tax revenues are a possible revenue source or it is possible for the regulator to tax the products of the incumbent monopolist and its rivals, it may be efficient to use general tax revenues or tax the products of the monopolist and rivals alike, with implications for access prices.
- If outside sources of funding are unavailable, prices on at least some services must be raised above marginal cost to allow the firm to just recover its total costs (and no more). If it is efficient to price one or more retail services above marginal cost in order to collect revenue it is usually efficient to use various forms of price discrimination (such as two-part pricing or

Ramsey pricing) to minimise the distortion in consumption that arises from pricing above marginal cost. When the different services produced by the regulated firm are substitutes or complements for one another, the regulator is not free to set these prices independently. If it is efficient to price a retail service above marginal cost, the price of substitute retail services should also be above marginal cost to maintain the correct relative retail prices – otherwise consumers will reduce revenue collection by substituting between the retail services. The same result also applies to the prices of inputs sold to rivals. In particular, where the regulator sets both the end-user prices of the incumbent and the access prices, if the final services provided by the incumbent and its rivals are substitutes and the incumbent's price is efficiently set above marginal cost, then either the access price should also be above marginal cost or the retail service taxed in order to preserve the correct relative final prices of the services of the incumbent and its rivals. In the case of perfect substitutes, the relationship between the access price and the retail price (or the portion of the retail price ultimately retained by the rival if the retail price includes a tax) should be given by the so-called Efficient Component Pricing Rule.

- Price discrimination at the retail level is undermined by competition unless supported by differentiated access charges or retail taxes. For example, in the case where the end-user services provided by rivals are a perfect substitute for the end-user services of the incumbent, consumers will purchase from the company providing the cheapest service. This can be a benefit for the regulator – if the structure of end-user prices of the regulated firm is currently inefficient, the regulator can set an efficient structure of access prices and rely on competition to force the end-user prices of the incumbent into line. It can also create problems for the regulator – if the discriminatory structure of end-user prices of the regulated firm is efficient, the undifferentiated access charges will undermine efficient end-user charges. Thus any price discrimination efficiently required in the incumbent's final prices should also be reflected in the corresponding access prices or, instead, supported by explicit taxes. At times, setting undifferentiated access charges may have undesired consequences: if the discriminatory structure of end user charges is the achievement of effective political constituencies (*e.g.*, geographic price averaging), the beneficiaries of the legacy rate structure may try to limit competition.
- In some cases it may be possible to use capacity-based pricing – a form of two-part pricing where the fixed part is proportional to the maximum capacity available. Long-term contracts at capacity-based prices have the effect of dividing the natural monopoly up into smaller parts which may compete with each other. In some cases it may be possible to delegate the regulated firm the responsibility to set its own regulated prices subject to a

ceiling on a weighted-average basket of prices known as a "price cap". This practice is common in the control of final prices. It is less common to regulate access prices through a price cap. Some economic theory suggests that it would be efficient to include both final prices and access prices within a single "global" price cap, although this has yet to be adopted in practice, in part due to concerns over the effect on competition.

- The interconnection of two telecommunications networks where both networks originate and terminate calls (whether fixed-fixed or fixed-mobile interconnection) is a two-way access problem. Most end-users are connected to only one network. That network therefore has a monopoly over calls terminating to those end-users. Under certain circumstances this network has the incentive to exploit this monopoly by raising termination charges. In order for the terminating network to be able to exploit its monopoly the following factors must be present: *a)* calling party pays; *b)* absence of reciprocity in termination charges (*i.e.*, the termination charges are not the same in each direction); *c)* consumers who care more about the price of calls they originate than the price of calls made to them; and *d)* retail charges which do not depend directly on the termination charge of the terminating network. This problem is sometimes known as the "terminating network monopoly" problem. How the regulator should set the termination charge in this context depends on balancing factors such as the effect of termination charges on final prices and on network penetration.
- Because of the terminating network monopoly problem, in the case of the interconnection of two reciprocal networks, regulators typically insist that the termination charges be reciprocal (*i.e.*, the same, or the same mark-up, in each direction). The effect of this requirement depends on the degree of price discrimination allowed in final prices. Where final prices are simple linear prices the interconnecting firms prefer to set high reciprocal termination charges to induce final prices at the monopoly level. Where final prices are simple two-part tariffs, and the networks are symmetric in size the networks are indifferent to the level of the termination charge. Where the networks are also allowed to discriminate between on and off-network calls, the firms will prefer to have low termination charges, to soften competition between the networks for final customers.
- The rationale for final price discrimination in this discussion is the efficient expansion of output in the presence of fixed costs (or other economies), beyond the level permitted by pricing at average cost. The pattern of actual final price discrimination found currently in regulated industries may not be efficiency enhancing. Access prices that support inefficient final prices may reduce efficiency.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Chapter 2 – The practice of access pricing

- Chapter 2 makes an attempt to relate the theory of Chapter 1 with the practice of access pricing in the telecommunications industry in OECD countries. Interconnection of two fixed-wire telecommunications networks providing point-to-point communications services requires access to either or both the local loops at the originating and terminating end of a call. When access is required at both ends, this is a one-way access pricing problem. When access is required at only one end, the correct model is the two-way access pricing problem. It is possible to distinguish four relevant access prices, corresponding to the charges for one-way and two-way call origination and termination. Of these, regulators usually directly set up to three – the one-way call origination and termination charges and the two-way termination charge (which is usually reciprocal). The fourth price, the price for two-way call origination, is not directly set by the regulator but is the difference between the retail price for a local call and the two-way termination charge (*i.e.*, it is the share of the retail price for a local call that is kept by the originating operator).
- Although the theory suggests that these four charges could be (and perhaps should be) set at different levels, when these four separate charges are set at different levels there is an incentive for operators to seek to arbitrage the different prices by re-routing traffic or re-organising the way they operate. For example, if two-way call termination charges are cheaper than one-way call termination charges, long-distance telecommunications operators have an incentive to disguise their traffic as two-way traffic to benefit from the lower rate (and *vice versa*). It may be possible to keep this traffic separate, but at the cost of greater regulatory intervention. Similarly, problems can arise if the two-way call origination charges are higher than the two-way call termination charges. Where this is the case, competing local carriers have an incentive to compete vigorously for end-users which terminate more calls than they originate, leading to an artificial distortion in competition and significant imbalances in flows of traffic across networks.
- Broadly speaking, OECD countries fall into two categories: The first category, which includes most European countries, includes those countries that choose to link all four charges: one-way and two-way call origination and termination charges. These countries set the same access prices for one-way and two-way traffic and for long-distance and local traffic. In addition, in many of these countries, the effective charge for originating a local call is close or equal to the charge for termination. These countries have chosen regulatory simplicity over the benefits of greater flexibility in setting access charges. As incumbent operators have introduced new innovative retail tariffs (such as flat-rate internet access) these countries

have had to respond by creating new classes of access services. On the other hand, those countries with unmetered local calling (US, Canada, Mexico, Australia and New Zealand) have generally not chosen to link these four access charges. These countries have, in principle, greater flexibility in setting access charges but have, in practice, faced the need for more active regulatory intervention to address the regulatory and competition problems that arise.

- Abstracting from the levels of prices, the theory highlighted the importance of reflecting the structure of access prices in retail prices and *vice versa*. In practice, there are both important similarities in the structure of retail prices and access prices and very important differences. For a large group of countries local, long-distance and international calling charges are primarily based on per minute charges (sometimes with a call set-up or "flag fall" fee). This is typically mirrored in the structure of call origination and termination charges, which are primarily based on per minute charges, often supplemented by a call set-up or flag-fall charge. In addition since, in most of these countries, retail charges are differentiated peak/off-peak, are geographically averaged and do not differentiate between business and residential customers, call origination and termination charges also are differentiated by peak/off-peak, are geographically averaged and do not differentiate between business and residential customers. For this group of countries there is at present a close match in the structure of retail prices and voice call origination and termination charges. However, some aspects of the price structures (*e.g.*, relatively high per minute charges on price sensitive demands) may inefficiently reduce usage and customer satisfaction. Incumbent operators are increasingly introducing innovative rate schedules such as "capped" (unmetered or flat rate) long-distance calls that may more closely match underlying costs but which lead to a mismatch in the structure of retail prices and access prices. Such schedules have given rise to competition complaints in Australia and New Zealand and are likely to give rise to increasing competition complaints in the future.
- In countries that have unmetered local calling, the use of per-minute access charges creates a clear incompatibility in the structure of access charges and local retail charges. These countries typically do not have competition for local calls using both call origination and termination. Australia, which has only a per-call fee for local calls, allows such competition by distinguishing a separate access charge for local calls based on a single per-call fee. In addition, in these countries, retail prices often differentiate between business and residential customers (with businesses paying higher rates). If access charges are undifferentiated, or are differentiated geographically (as in Australia) new entry may focus exclusively on the business sector and in low cost areas and may force retail rate restructuring. A system of explicit taxes and subsidies

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

(*e.g.*, "universal service funds") or appropriately chosen access charges may improve potential rivals' entry and investment incentives in a competitive environment.

- A minority of telecommunications access services are set using a "retail-minus" approach. Under this approach the access price for a given access service is equal to the retail price for the corresponding retail service less a discount. If the size of the discount is chosen correctly both the structure and level of retail prices is reflected in access prices. Even in the case of those services for which a retail-minus approach is not directly used, most countries do not completely ignore the link between retail prices and access prices. For example, several countries explicitly mentioned that they take into account the relative levels of the access charges and retail charges when approving tariffs.
- Most OECD countries have special access arrangements for handling calls to Internet service providers ("ISPs"). The incentive for competing local carriers to compete for the business of ISPs depends on whether it is more profitable for them to target end-users that have a net outflow of calls (*i.e.*, originate more calls than they terminate) or end-users which have a net inflow of calls (*i.e.*, terminate more calls than they originate). If carriers receive more net revenue from call termination than they do from call origination there is a strong incentive for them to target ISPs – which have a strong net inflow of calls. This is particularly a problem where local calls are free because, in that case, the net revenue from call origination is zero or negative. In the US, CLECs have actively targeted end-users with a net inflow of calls. In fact, CLECs terminate on average eighteen times more call minutes than they originate. Problems can also arise from a mismatch in the structure of retail prices for Internet service and access prices. The introduction of flat-rate Internet service has given rise to competition concerns in many countries. Most OECD countries, starting with the UK, have chosen to respond by reflecting the structure of retail prices in access prices, by requiring a corresponding flat-rate internet access call origination service ("FRIACO").
- Virtually all OECD countries mandate access to (known as "unbundling" of) copper-wire local loops. These loops can be used to provide two distinct services – conventional low-bandwidth voice services and high-bandwidth Digital Subscriber Line ("DSL") services. Theory suggests that, if the rights to provide low- and high-bandwidth services on a loop are sold separately, the price (or prices) of the unbundled local loop should depend on the use (or uses) to which the local loop is put. In most cases the retail price for ADSL service is geographically averaged and does not differentiate between business and residential customers. Correspondingly, in most countries the access price for unbundled local loop ("ULL") is typically geographically averaged and does not

differentiate between business and residential customers. Australia, which has geographically averaged retail prices, has chosen to geographically de-average ULL prices. This increases the likelihood of both intense competition in CBD areas and limited entry in higher cost areas unless retail prices are deaveraged or explicit taxes and subsidies are used to support the retail price structure.

- The terminating network monopoly problem has arisen in some countries, most notably in the US, in the interconnection of CLECs and IXCs. The traffic from IXCs terminating in CLECs is all one-way and reciprocity does not apply. CLECs exploited their monopoly by setting termination charges above the charges of the ILEC in the same area. In those countries with Calling Party Pays ("CPP") for mobile calls, fixed-to-mobile calls fulfil the requirements for a terminating network monopoly problem – each mobile network has a monopoly over calls to its subscribers, reciprocity does not apply in fixed-to-mobile calls (termination on fixed networks usually has a much lower charge than termination on mobile networks) and fixed-to-mobile retail charges seldom depend directly on the identity of the terminating mobile network. As long as each mobile subscriber cares only about the calls he or she makes, competitive pressure on mobile termination charges is weak or non-existent. In this circumstance, the relevant market for call termination is the market for the termination of calls to each individual mobile subscriber. Each terminating network has a monopoly in this market, even if the terminating network is very small. The UK has recently extended its regulation to call termination on all four UK mobile networks, using an RPI-12% approach to reducing mobile termination charges. Australia has chosen to tie the downward movement of mobile termination charges to the downward movement of outgoing mobile call charges.

Chapter 3 – The measurement of costs of access services

- Since many countries choose (or are required) to set access charges on the basis of the underlying cost, it is essential to understand what it means for regulated prices to be "cost based". This chapter begins with a review of the theory on how to allocate the cost of sunk investments over time. Abstracting from issues of uncertainty and risk aversion a regulated firm is assured of just covering its costs if and only if the net present value of its total net payments to investors (*i.e.*, the revenue earned less the capital expenditure and operating expenditure each period) over its life is zero. The excess of revenue over expenditure in any one period can be high or low but over the whole life of the firm, if the firm is not to earn excess returns, the net present value of the payments to investors must be zero. Prices can be said to be based on "historical cost" if they satisfy the constraint that the net present value of the payments to investors over the life of the firm is zero.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

- Future events that affect the value of capital assets can be uncertain. Uncertainty is resolved over time and, as it is resolved, the best way of utilizing existing assets change. A benefit of competitive markets is that such markets efficiently incorporate both uncertain and revealed information in asset prices. The amount that was paid for an asset, or its value at a previous time, does not directly affect its value today, as each homeowner knows. Regulators also may choose to use pricing methodologies that incorporate new information that is revealed during the lifetime of an asset. A regulatory methodology for setting prices and adjusting them in response to new information must satisfy a broad but important constraint: an efficient investment must yield an "expected (net present) value" of payments to investors over the life of the firm equal to zero. If less than zero, the regulated firm will not voluntarily undertake the investment. If greater than zero, the incentive to invest is too high and the incentive to consume too low. This is a notion of "cost-based" pricing that is more general than the previous one, allowing a range of pricing methods that provide efficiency benefits by incorporating new information.
- "Forward-looking cost" pricing methodologies suitable for regulation are a subset of these methods. Long- and short-run forward-looking cost methods periodically adjust prices on the basis of the current value of capital assets and the current expectation of depreciation. Long-run forward-looking methods value capital assets by the current costs of duplicating the functions of such assets in today's markets. Short-run methods, as more often used in electric markets, consider the relationship between the amount and type of existing (*i.e.*, sunk) capital stock and demand in making this valuation. These methods essentially "mark to market" at each adjustment so that, whenever prices are set, the resulting expected net present value of future payments equals the current value of the asset, plus, if there are scale economies, the additional amount needed to cover the difference between expected average and marginal cost. Past prices or payments to the monopolist do not directly affect the value of forward-looking cost prices. Thus, purely forward-looking cost price methods would require modification to allow for any anticipated revenue transfers across time.
- It is conventional for regulators to approach the task of historical cost ratemaking by defining an amount called the firm's "asset base". This amount is increased each period by the size of any new investment and decreased by an arbitrary amount known as "depreciation". The total allowed earnings in the period are conventionally given as the sum of the "return on capital" (equal to the allowed cost of capital times the asset base) and the "return of capital" (equal to the depreciation). When the regulator's task is expressed in this way, and when the scrap or salvage value of the firm's assets is zero, any path of depreciation which sums to the total

capital expenditure of the firm corresponds to a path of payments to investors with a net present value of zero. Rather than focus on controlling the firm's earnings, the regulator can instead focus on choosing a path for depreciation. Alternatively, the regulator could focus directly on the path of the firm's asset base. Any path of the asset base which reduces to zero over the life of the firm corresponds to a path of payments to investors with a net present value of zero.

- Many regulatory authorities choose to focus on the choice of a path of depreciation. It is common to use a simple formula for depreciation such as "straight-line" depreciation or "sum-of-the-years'-digits". Since, under this approach, the asset base depends only on the historic or original cost of the underlying asset, this approach can be called the "historic cost" or "original cost" approach. The corresponding path of earnings is declining in periods between new investment. If the prices of new assets are constant or increasing, the resulting path of earnings may have a large upward jump when major new investment is required, which may give rise to "rate shock".
- Choosing an efficient cost allocation (*i.e.*, an efficient path of depreciation or path of the asset base) requires consideration of the effect of prices on entry or consumption behaviour. In industries where replacement costs are declining (such as the telecommunications industry) concerns are often raised about the potential for inefficient entry. When the replacement cost of a firm's assets are decreasing, if prices (and earnings) remain constant there is a risk that the replacement cost will drop to the point where another firm will prefer to duplicate the asset, which may be inefficient, rather than purchase from the regulated firm. This can be prevented by choosing a path of earnings (and probably prices) that is also declining. In some cases this constraint implicitly determines the required path of earnings and therefore the path of depreciation or the asset base. Where the replacement cost is declining at a constant rate the path of earnings which prevents inefficient duplication is also declining at the same constant rate and (therefore) is a simple constant fraction of the replacement cost.
- Rather than focus on the effect on entry, it is also possible to find an efficient cost allocation through analysis of the effect of prices on consumption behaviour. The optimal path of earnings is the path for which the resulting end-user prices maximise consumer welfare subject to the constraint that the regulated firm receives sufficient revenue to cover its investment costs. This approach leads to the familiar Ramsey formula, which in its simplest form states that prices should be higher in periods with less elastic demand. In addition, many of the other familiar results from the regulation of natural monopolies also apply – including the benefits of two-part pricing, peak-load pricing, price discrimination and so on. The failure to choose a cost allocation under which prices vary with demand conditions not only lowers

total welfare, but also has the effect of deterring some efficient investment. For example, a requirement that prices should cover average costs in all demand conditions forces prices to be higher when demand is weaker – reducing consumers' surplus and thereby deterring efficient investment. For the same reason insisting on a constant price when demand is uncertain may deter investment.

- One common approach is to allow the asset base of the regulated firm to vary in accordance with the (depreciated) cost of purchasing or constructing a modern equivalent asset. This depreciated replacement cost (or "DORC") approach leads to problems when new investment is "lumpy" (*i.e.*, when there are economies of scale in investment). In this case, the cost of providing the current level of network services depends in part on the history of demand. For example, it may have been efficient to add capacity in increments even though the resulting costs exceeds the cost of creating an asset of sufficient size to meet demand today. In this context, insisting on the DORC methodology may require undesirable fluctuations in earnings or may leave the regulated firm undercompensated. This may explain why most regulators, although they state they are using a DORC methodology, adopt a scorched node rather than a scorched earth approach to valuing telecommunications assets. In addition, when the payments to providers of capital depend only on replacement costs, the regulated firm may not undertake certain capital expenditure (such as "refurbishment") which is necessary but which does not affect replacement cost.
- Most of the discussion in this chapter concerns a situation with no uncertainty about the future. In practice, investment decisions in the real world are fundamentally affected by uncertainty about future outcomes. It is possible to analyse the effect of certain types of uncertainty. If the uncertainty in earnings is stationary it is possible to incorporate the effect of uncertainty in the form of an adjustment to the cost of capital allowed to the regulated firm. Specifically, when the uncertainty in earnings is stationary and normally distributed the interest rate should be given by the conventional weighted-average cost of capital formula. Typically the cost of equity (and sometimes debt) is given by the traditional capital asset pricing model. Depreciation should be calculated on the basis of the expected (or average) earnings given the regulated price (not the actual earnings *ex post*).
- In some cases the uncertainty in earnings is not stationary – that is, the resolution of the uncertainty in one period affects future periods. In this case it is efficient for the path of allowed earnings (and therefore depreciation and the asset base) to be allowed to depend on new information that is revealed over time. For example, where there is a risk that a technological development will prematurely reduce demand for the regulated firm's services this risk should be reflected in higher allowed earnings to the regulated firm than would be

given by the simple models. If the regulated firm is to be adequately compensated on average then it must be that the asset base at the end of each regulatory period should be equal to the average or expected value of the (revised) asset base at the start of the next period.

This report starts from the underlying assumption that economic theory provides useful insights which can assist policy-makers and regulators as they seek to develop policies regarding access prices. However, there are limits to the extent which theory can guide practice. Some of these factors are set out in Box 1. Economic theory relies on a set of strong assumptions and regulators can not be assured to what extent the economic theory assumptions or even the framework apply.

Note that this report focuses on how regulators should set the *prices* of essential inputs. We will not focus on other aspects of access regulation, such as control of the quality or the capacity of the access service, whether and which inputs to unbundle, or the incentives and ability of regulators to carry out each pricing method. Nor will we discuss broader public policy approaches towards vertically-integrated natural monopolies, such as vertical structural separation. A substantial discussion of the essential facilities problem with a discussion of different policy approaches for promoting competition can be found in our previous study on structural separation in public utility industries, OECD (2001d). The efficiency of the access pricing methods discussed in this report often is conditional on the values of retail prices. Sometimes retail prices are not fixed. In industries that are more naturally competitive, such as airlines and trucking, dissatisfaction with regulated rates has been a significant reason for abolishing the regulator or deregulating. Certain access pricing methodologies may have value because they are likely to make retail prices more efficient by limiting the retail pricing discretion of the regulator. These aspects of access pricing often involve political economy and are not discussed in detail in this report.

This report was prepared in the period October 2001 – August 2003. In the preparation of this report OECD countries were invited to make submissions describing their regulatory regime for access prices and their processes for measuring costs. Unless otherwise indicated, citations in this report are to those submissions. The glossary at the end of the report provides definitions for some telecommunications-specific terminology and abbreviations.

Box 1. **The relationship between theory and practice**

Economic theory is clearly a primary input into the decision making processes of policy makers and regulators. But, as this report shows, economic theory cannot always be perfectly reconciled with practice. Does this imply that the economic theory is "wrong"? Or is it that the practice is "wrong" and should be improved? Or neither?

There are many reasons why theory and practice may diverge. Any theory is based on a set of assumptions (or "models"). These simplifying assumptions may not accurately or adequately reflect the most important aspects of reality. In particular:

- An economic theory may not adequately reflect differences in the information that is available to the relevant parties and the cost of independently verifying that information. It has been widely recognised in economics for at least a couple of decades that different access to information lies at the heart of the regulator's problem.
- An economic theory must make assumptions about the way that firms behave or the way that regulators behave which may not reflect important aspects of reality, such as ignoring important political pressures under which regulators' operate.
- An economic theory might ignore costs which are important in practice, such as the cost of acquiring the necessary information to implement the recommended approach or the cost of small deviations from the recommended approach (*i.e.*, the "robustness" of the solution).
- An economic theory might not adequately reflect constraints which are important in practice such as the limited time or resources that a regulator might have to reach a decision (*e.g.*, staff resources, or computational resources) or the scope for judicial appeal or review.
- An adequate economic theory which captures all of the most important factors might simply not yet exist.*

Prudent use of economic theory requires that these limitations and caveats be recognised and addressed wherever possible. Theory is an important guide to policy makers but there may still remain legitimate reasons for differences between theory and practice.

* Economic theory will sometimes "lag" and sometimes "lead" practice.

ISBN 92-64-10592-1
Access Pricing in Telecommunications

Chapter 1

The Theory of Access Pricing

1. Introduction

This chapter seeks to explain as clearly as possible the economic theory of access pricing, illustrating the theory with applications in the telecommunications sector. This chapter draws heavily on a survey paper by Armstrong (2001) prepared for the Handbook of Telecommunications Economics.

1.1. The basic access problem and a word on terminology

The basic access problem arises when the provision of a complete service to a final customer requires the use of two or more complementary services, one or more of which is non-competitive (*i.e.*, cannot sustain competition). It is easy to find examples of complementary services – examples include aircraft engines and airframes, glass bottles and wine, left shoes and right shoes – in fact virtually any two vital inputs into a production process are complementary. Complementary services abound in telecommunications – the provision of an end-to-end telecommunications service (such as a telephone call or an Internet connection) almost always requires a combination of a number of separate components such as call origination, transportation and call termination.

Throughout this report, unless otherwise stated, we will assume that the two complementary services are consumed in fixed proportions. We will assume that the final consumer or end-user always consumes precisely one unit of the competitive and one unit of the non-competitive service at a time, yielding one unit of final output. In this context, the consumer is indifferent to purchasing *a)* both complementary services separately from two different firms; or *b)* both services together from the provider of the competitive service (who purchases the non-competitive service as an input); or *c)* both services together from the provider of the non-competitive service (who purchases the competitive service as an input). In every case the consumer cares only about the total price and quality of the combined service.

For example, suppose an end-user wishes to purchase dial-up Internet services. Assume this service requires the combination of just two inputs in fixed proportions – a local loop connection from the location of the user to the computer of an Internet Service Provider ("ISP") and a computer providing Internet services. The relationship between these two inputs and the end-user can be organised in three ways: *a)* the end-user could purchase delivered internet services from the local loop provider which, in turn purchases internet services from the ISP; *b)* the end-user could purchase delivered internet services from the ISP, which, in turn, purchases call origination services from the local loop provider; *c)* the end-user could separately purchase local loop services from the local loop provider and internet services from the ISP. If the relationship between the two inputs and the delivered service is the same in each case, all three of these forms of organisation are economically equivalent. All can be found in OECD countries (and often simultaneously in the same country).[1]

The assumption of fixed proportions is not trivial and is not always met in practice. If it is met, and the monopoly input cannot be bypassed, then charging an access price that is above marginal cost is equivalent – with respect to the amount of revenue raised, the demand for access, and the demand for the final product – to placing a tax of the same amount on the final product and pricing access at marginal cost. Neither the mark-up on access nor the tax distorts input proportions. Since we normally make the assumption of fixed proportions in this report, we use the term "access price" to refer the marginal cost of access plus any premium whether in the form a mark-up on access or a tax on final output. Where input proportions are not fixed ("variable"), the efficiency with which the final service is produced by entrants that use monopoly access declines as the price of the monopoly input is raised above marginal cost: Competitive producers of the combined service respond by substituting more of the competitively provided input to produce any given amount of output. With variable input proportions the equivalence between mark-ups on inputs and taxes on outputs does not hold. The non-competitive input must be priced at marginal cost and a tax placed on the final service in order to raise revenue without distorting the efficient mix of inputs.

When describing an access problem it is common to use terminology that implies that one firm is "upstream" from another firm. Although sometimes helpful, such terminology is often misleading. As the above examples make clear, when dealing with complementary services, the designation as to which firm is the supplier of services and which firm is the purchaser can be entirely arbitrary. It often has no economic significance. For example, in the example of call origination for Internet services – is the provider of local loop services upstream or downstream from the Internet service provider? Does it matter?

For the sake of clarity, this report will often continue to use the terms "upstream" and "downstream". However, it is useful to keep in mind that these expressions are ambiguous. All that matters is the price paid by the end-user and how that revenue is divided between the firms providing complementary services.

1.2. Different categories of access pricing problems

It is possible to distinguish different categories of access pricing problems, depending on which firms must purchase essential inputs from which other firms. In this report we will distinguish the following three categories of access pricing problems:

1. The "one-way" access pricing problem, in which, in order to provide a complete service to end-users, rival firms or networks must purchase essential inputs from a monopoly firm or network but not *vice versa* (the monopoly firm can provide a complete service to end-users without any inputs provided by its rivals). This is the "classic" access pricing problem. Most of the conventional access and essential facility problems in public utility industries can be analysed within this framework.
2. A "two-way" access pricing problem in which, in order to provide a complete service to end-users, competing firms or networks must purchase essential inputs from a monopoly firm and, in addition, that monopoly firm must purchase inputs from the competing firms (but the competing firms do not purchase inputs from each other). This model is applied in this chapter to the interconnection of competing mobile networks with a fixed telecommunications network.
3. A "two-way" access pricing problem in which, in order to provide a complete service to end-users, both firms must purchase essential inputs from each other. This model is used in this chapter in the discussion of interconnection of two competing fixed or mobile telecommunications networks.

Economic models of access pricing (especially two-way access pricing) tend to become relatively complex rather quickly. No general model of two-way access pricing has yet been developed. The models that we do have tend to be rather specialised and it is difficult to derive general principles. It is common to make rather restrictive assumptions in order to keep the analysis manageable. Nevertheless, certain general principles can be enunciated for the case of one-way access pricing. And, as we will see, in the case of two-way access pricing, the range of models that have been developed to date present an interesting picture of the outcomes that can emerge.

This chapter (and, indeed, this report) will not focus on the issue of asymmetric information. That is, we will assume that the regulator has adequate information on the cost structure of the regulated firms. This allows us to focus on the issue of how this information on the cost structure should be used to set the regulated access prices. Of course, in practice, the regulator may not have this information. Taking into account the problem of asymmetric information may, in certain circumstances lead to quite different results. This caveat should be kept in mind.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

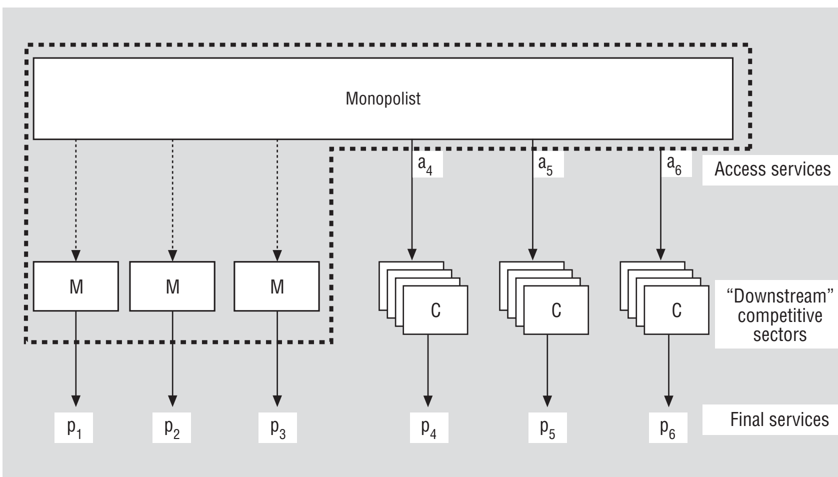
2. One-way access pricing

The "classic" or "traditional" access pricing problem arises when the provider of non-competitive services sells inputs to the providers of the competitive services, which are then combined and sold to final consumers or end-users. We will assume that the provider of non-competitive services does not need to purchase services from the providers of competitive services – this is the conventional "one-way" access problem.

For simplicity we will call the provider of non-competitive services the "monopolist". We will refer to the non-competitive services sold to the downstream competitive firms as "access services". The services that are sold directly to final customers are "final services".

The monopolist may itself provide final services in competition with the providers of competitive services. This market structure is illustrated in Figure 1.

Figure 1. **Market structure of a one-way access problem**



Source: OECD.

The principles governing access pricing in this context are little more than an application of the principles governing the efficient pricing of a natural monopoly. This section therefore begins with a review of those principles. The second part of this section applies those principles to the specific context of access pricing.

2.1. Review of the principles for efficient regulation of a natural monopoly

Suppose we have a monopoly firm with a known cost structure that produces a range of services. The demand for any one of these services depends on both the price charged for that service and the prices charged for any other complementary or substitute services. The problem faced by a benevolent regulator is to find a set of prices for these services which maximise overall economic welfare[2] subject to various constraints.

The efficient set of prices in any given circumstance depends on a variety of factors, including, importantly, the set of *instruments* and *information* available to the regulator and the *constraints* which the regulated prices must satisfy. More specifically, the efficient set of prices will depend on factors such as how finely the regulator (or the regulated firm) can differentiate its prices to different final customers and whether or not the regulator can use other public policy tools such as general or industry-specific taxes and subsidies. The more finely differentiated a set of prices are, and the more instruments a regulator has as its disposal, the more efficient the final outcome will be. The most efficient set of prices attainable by the regulator will also depend on such constraints as whether or not the monopolist is required to break even (*i.e.*, earn revenues sufficient to cover costs) or whether prices are required to be geographically uniform or held below "cost" for certain groups.

The first key principle in this context is well known – when the regulator is able to choose all the prices of the monopolist simultaneously, without additional constraints, total welfare is maximised by setting all prices equal to marginal cost. This result is quite intuitive. With simple linear pricing, consumers will consume up to the point where the marginal utility from the last unit consumed is equal to the price. Overall welfare is maximised when consumers consume up to the point where the marginal utility from the last unit consumed is just equal to marginal cost. Therefore the regulator should set the price of each good equal to marginal cost.

> ***Principle 1: (Marginal cost pricing)*** *If all of the prices of the monopolist are variable and if the revenue obtained by the monopolist when prices are set equal to marginal cost is sufficient to cover the total costs of the monopolist, or if the deficit between the total revenue and the total costs of the monopolist can be recovered in some other way which does reduce economic efficiency, the efficient price for each service is simply equal to the marginal cost of that service.*

In circumstances where marginal cost pricing is possible, it is interesting to note that the pursuit of economic efficiency gives rise to no further need for consideration of the nature of demand for a particular service; nor is there any need to take into account the effect of prices on substitution between different services. This is the one limited context in which prices are determined entirely by "costs" and therefore the only context in which the notion of "cost-based" prices is unambiguous.[3]

In practice, it is often the case that prices for certain services of public utility industries are chosen to collect more money than pricing at marginal

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

cost would raise. This can be due to scale or other economies which result in average cost exceeding marginal cost, or due to the use of these prices to fund other public policy objectives[4] (such as geographic uniformity of tariffs or "universal service"). If some prices are moved away from marginal cost, does that have any implication for the efficient choice of the remaining prices?

The answer is yes: Where the goal is to efficiently raise a given amount of revenue from the industry, if it is efficient to price a retail service above marginal cost, then (other things equal) substitute retail services also should be priced above marginal cost. If these relative prices were not maintained, consumers would switch purchases to the lower-priced substitutes, with two alternative consequences: First, if net revenue must be preserved, greater departures from marginal cost pricing among the remaining services would be required, reducing overall economic efficiency. Second, the regulator or other policy maker responsible for the pricing constraint may abandon its attempts to achieve its revenue goal through industry pricing. Since the goal may or may not be "costly" – a constraint on economic efficiency – policies that lead to abandoning the goal may or may not be economically efficient.[5]

A corollary is that two services which are perfect substitutes must have the same price – if one is above marginal cost to pursue a public policy objective, the other must be as well, or the public policy objective will not be achieved.[6] These results are shown more formally in the appendix to this chapter. As we will see later, these effects become important in the context of access pricing, when substitution away from one service towards another implies "distorting the playing field" between competing companies.

Principle 2: (Substitution effects between services) *If some of the prices of the services of the monopolist are fixed (*e.g.*, if some group of prices are set through some other process) and are set above or below marginal cost, the efficient price for any services which are complements or substitutes for the fixed-price services must also be set above or below marginal cost (in the same direction for substitutes and in the opposite direction for complements) to raise a given amount of revenue efficiently.*[7]

In most public utility industries, pricing at marginal cost would prevent the regulated firm from earning enough revenue to cover its costs. There is a substantial debate in some industries (such as the railroad industry) whether the prices for access services should cover total cost rather than marginal cost. There are economic arguments for and against full cost recovery.[8] We will not go into these arguments here. We will merely note that if it is essential to raise prices above marginal cost to recover at least some proportion of the total cost, the efficient price levels are given by the familiar Ramsey price formulas. (See example 1 in Box 1.)

***Principle 3: (Ramsey pricing)** If all the prices of the monopolist are variable and if raising funds from outside the sector to cover the deficit between revenues and costs is not possible, the prices of the monopolist must be raised above marginal cost to a level which allows the recovery of (at least some of the) total costs. The amount that individual prices should be raised depends on the (super-) elasticity of demand for those services. Services with a lower elasticity of demand will tend to have higher prices. At the same time, to maintain efficient substitution between the services, the relative position of services which are complements and substitutes should be maintained – if the price of a service is raised, the price of all substitute services should be also raised and the price of all complementary services should be lowered.*

To the extent that Ramsey pricing involves pricing above marginal cost, it involves a loss of allocative efficiency – customers who have a willingness to pay above the marginal cost but below the price are inefficiently dissuaded from making a purchase. This efficiency loss can be reduced by techniques that more closely match a customer's willingness to pay for a good to the price they are charged. In particular, through various forms of price discrimination – *i.e.*, charging different prices to different customers, or charging prices that are differentiated geographically, or by time, or by the number of units purchased. From a theoretical perspective, price discrimination is equivalent to expanding the range of services that are sold. The prices for those individual services are then set according to the principles above including, in particular, the principle of Ramsey pricing. See example 2 in Box 1.

For example, rather than charge a single price (equal to average cost) for a service, it is often more efficient, if feasible, to charge a two-part price, with a fixed charge for "access" (not to be confused with access to essential inputs) and a separate charge for "usage". Ramsey pricing can then be used to set the price for each of these two new services. It is important to take into account the fact that these two services will usually be complements – a reduction in the usage charge may increase the number of customers wishing to sign up for service. Depending on the relative elasticities, it may, therefore, make sense to charge above marginal cost for usage in order to lower the fixed charge, to encourage greater demand for the overall service.

One variation of two-part pricing that is worth noting is capacity-based pricing. A capacity based price is similar to a standard two-part price except that the size of the fixed part of the two-part tariff determines the maximum quantity that the customer may consume. Capacity-based pricing can be thought of as a three-part charge with a fixed component (which determines the capacity) a usage component for a quantity purchased less than or equal to the allowed capacity and then a very high price for quantity purchases above that capacity limit. Capacity based pricing is discussed further below.

***Principle 4: (Price discrimination)** Assume it is necessary to raise greater revenue from utility services than can be collected by pricing these services at marginal costs. If it is possible to discriminate in the setting of prices –* i.e.*, if it is possible to charge different prices to different customers, or for services purchased at different times of the day, or at different geographic locations, or if it is possible to charge a price which depends on the number of units purchased – then it is usually efficient to do so. In other words, it is generally efficient to use two-part or multi-part pricing, or to use second and third-degree price discrimination. Such discrimination is theoretically equivalent to expanding the range of services. The same formulas for Ramsey pricing apply to the pricing of the new expanded range of services.*

Price discrimination is only possible when the regulator (or the regulated firm) has the information necessary to differentiate the prices (*i.e.*, when the regulator or the firm can identify the customer, the time of day or geographic location, or the number of units purchased), and when the regulator or the firm can prevent resale to other customers.

Even where price discrimination of a service is feasible, it may be hindered by the absence of price discrimination in substitute or complementary services. Price discrimination operates by dividing existing services into new services with differing elasticities of demand, and offering higher prices to those new services with lower elasticities of demand. However if a substitute product is available to all the end-users (because the price of this product is not or cannot be differentiated) customers facing higher prices will tend to substitute towards the undifferentiated product. In fact, the regulated firm may prefer to withdraw a substitute product from sale when it hinders its ability to price discriminate in another service. See example 3 in Box 1.

These observations may be important in the context of access pricing because even if price discrimination is possible for a final service of the monopolist it may not always be possible for access services which are used to produce substitutes for the monopolist's final service.

***Principle 5: (Price discrimination and substitute services)** The extent to which the regulated firm can use price discrimination to differentiate the prices of its services is limited by the extent to which price discrimination is feasible not only in the services themselves, but also on any substitute services. When two services are close substitutes and price discrimination is possible on one but not the other, the efficient pricing structure (and/or ensuring the viability of the monopolist) will involve raising the price or withdrawing from sale the service for which price discrimination is not possible.*

In some circumstances, if the monopolist's prices are high enough it may be possible for a firm (or a group of existing customers of the monopolist) to produce a set of services itself rather than purchase the same services as inputs from the monopolist. If there are economies of scale such "new entry" into the market for the essential inputs involves inefficient duplication of the existing assets of the monopolist.[9] If new entry leads to higher costs, the potential for inefficient entry should be taken into account when setting efficient prices.

Specifically, the possibility of such entry effectively places a ceiling on the prices that can be charged for a group of services. In theory, new entry may be possible whenever the revenue from a set of services exceeds stand-alone cost.[10] This new requirement may conflict with the other principles we have highlighted above. It may not be possible to simultaneously satisfy the other principles above and to prevent entry. For example, "pure" Ramsey pricing may require that the price for a service exceed its stand-alone cost. It may not be possible both to impose efficient Ramsey pricing and to prevent new entry.

This conflict between objectives can be overcome by introducing a new instrument. Specifically, what is needed is an instrument that would break the link between the prices charged by the monopolist and the prices faced by consumers. If this were possible, the prices charged by the monopolist would be lowered to deter entry, while the prices actually faced by consumers would be maintained in order to preserve the correct relationship between the consumer prices. One such instrument is a tax on the consumption of a particular service. This tax would be imposed on all consumers whether they purchase from the monopolist or produce the service themselves. With this tax, the monopolist's prices could be lowered, so that the monopolist's price remains below stand-alone cost, but the effective price faced by the consumer (*i.e.*, the sum of the monopolist's price plus the tax) would be maintained at the previous level.

Principle 6: (Control of entry or bypass) *If it is possible for other firms to produce the essential inputs themselves (at a higher cost than the monopolist), this factor must be taken into account when setting the access prices. Since entry into the provision of a group of services is more likely the higher the revenue derived from those services, the possibility of entry places a constraint on the revenue earned from any group of services. The possibility of inefficient entry or bypass can be reduced or eliminated through the use of taxes or levies on consumers of those services (whether purchased from the monopolist or not). The access prices of the monopolist can then be reduced to a level that does not induce entry while the tax ensures that the total amount paid for the final services reflects the correct relative and absolute level.*

Ramsey pricing has often been derided as impractical. It is true that calculating the efficient set of Ramsey prices requires substantial amounts of information, such as information about how demand for each service responds both to changes in its own price and changes in the price of other services, together with detailed information about the structure of cost. The

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

regulated firm itself is likely to have better information about cost and demand than the regulator. The regulator can make use of the information held by the regulated firm by delegating responsibility for setting individual prices to the regulated firm subject to an overall constraint, such as a cap on the "average" price of a basket of services.

A few caveats on the use of price-caps are worth mentioning here. First, the price that a firm will choose will depend on the marginal cost as perceived by that firm. In most cases this does not pose a problem. But what if the price of an input purchased by the regulated firm exceeds its "true" marginal cost for some reason? We will see later that this situation might arise when the regulated firm needs to purchase inputs from other firms (in the case of two-way access issues) and those inputs are priced above their marginal cost. In this case the regulated firm will choose a price for its output which is "too high" from the perspective of overall welfare.

Second, the use of price caps, although substantially reducing the information requirement of the regulator, does not eliminate it entirely. If price caps are to operate efficiently, the regulator must still choose how much "weight" to assign each price in the calculation of the overall average price. In theory, these weights should be simply proportional to the quantities sold by the regulated firm at the efficient prices, but information on quantities actually sold is not available until after the fact. Choosing the weights in a price cap is a difficult, although not necessarily intractable, problem in both theory and practice.

One obstacle to the use of price caps is simply the number of these weights that may need to be chosen. As discussed above, the regulated firm may wish to price discriminate in its sales to its customers. Although this does not pose a theoretical problem to the use of a price cap (price discrimination simply expands the range of services that are to be included in the price cap), with price discrimination the number of distinct services can increase rapidly. The need to set the correct weights for each of these services could prove to be a sizeable obstacle for the regulator.

A more serious critique to the use of price caps is that if the regulator allows price discrimination, as mentioned earlier, and if price discrimination of substitute products is not also allowed, the regulated firm may simply choose to not sell the product for which discrimination is not possible. As we will see later, in the context of access pricing, a firm subject to a price cap may simply refuse to sell access.

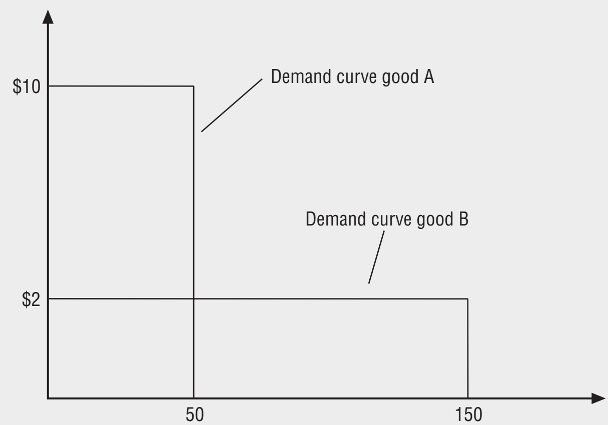
A final critique to note is that when the regulator is uncertain about demand for the services of the regulated firm and marginal cost is declining it may be preferable for the regulator to not delegate any pricing authority to the regulated firm (but simply to set a single regulated price).[11]

Box 2. **Price discrimination and natural monopoly regulation**

Example 1:

Consumers can be made better off by allowing the regulated firm to discriminate in the way it sets its prices. In some cases a regulated firm that is not allowed to discriminate may not be able to earn enough revenue to survive at all.

Suppose that a monopolist produces two goods at zero marginal cost, and a joint fixed cost of $500. Demand for each good is as follows – the maximum price that consumers are prepared to pay for good A is $10. At any price less than $10, consumers purchase 50 units. In the case of good B, the maximum price consumers are prepared to pay is $2 and will purchase exactly 150 units if the price is less than $2. This is illustrated in the diagram below. The problem is to find the set of prices that maximises consumers' surplus subject to the condition that the prices must yield enough revenue to enable the monopolist to cover its fixed costs of $500.



Source: OECD.

We could start by trying to use a common mark up over marginal cost (since marginal cost is zero in this example, this corresponds to a common price for the two goods). In this case, the only common price which yields revenue sufficient to cover fixed costs is the price of $10 – at this price only good A is sold (and none of good B) and the total revenue is 50 × 10 = $500. At this price consumers' surplus is zero and the monopolist earns just enough revenue to cover its fixed costs. There is no other common price that yields enough revenue to cover fixed costs.

Box 2. **Price discrimination and natural monopoly regulation** *(cont.)*

Is this the best that the monopolist can do? What if it lowered the price of good B to $2 and the price of good A to $4? In this case it would sell 50 units of good A and 150 units of good B, for a total revenue of $200 + $300 = $500. Consumers' surplus is now $300. This is the efficient outcome. Intuitively, by lowering the price on the more elastic good, the contribution to fixed costs from that good increased, allowing the price of other, less elastic goods, to be lowered.

What if the maximum price that consumers were prepared to pay for good A was only $9? In this case the monopolist would not be able to survive if it were forced to use a common mark up – the *only* way the monopolist could survive is by using some form of differentiated pricing between the two goods. Finally, it is worth noting that the regulator could induce the regulated firm to choose the efficient prices by allowing the regulated firm to choose its own prices subject to a price cap. Specifically, the regulator could allow the monopolist to choose any prices provided that the total revenue at those prices, assuming 50 units of good A and 150 units of good B, is less than $500.

Example 2:

Allowing price discrimination across customer groups can leave all customer groups better off.

Suppose that, as in the previous example, we have a monopolist who produces a good with zero marginal cost and a fixed cost of $500. But now, rather than assuming the monopolist produces two goods, let's suppose that it produces a single good but there are two classes of customers. There are 50 consumers in class A and 150 in class B. Class A consumers purchase one unit of the good if the price is below $10, and not otherwise. Class B consumers purchase one unit if the price is below $2 and not otherwise.

If the monopolist cannot distinguish between consumer classes then the only price that allows the monopolist to break even is $10. At this price only class A consumers purchase. No class B consumers purchase. As before, however, if the monopolist can distinguish between customers, it can lower both prices – the price for class B consumers could be lowered to $2, inducing them to consume more, and lowering the price of class A consumers to $4. (In fact, any pair of prices is efficient as long as the price to class B consumers is less than or equal to 2 and the price to class A customers is equal to 10 less three times the price to class B consumers.)

We could also change the problem so that class A and B consumers also differ in the quantity they consume. Suppose, for example that there are only 50 consumers in class B, but they consume 3 units each if the (marginal) price drops below $2. In this case, the monopolist can also discriminate between these

Box 2. **Price discrimination and natural monopoly regulation** *(cont.)*

two consumer classes with a non linear price. For example, the monopolist could charge $4 for the first unit consumed and then $1 thereafter. Another, equivalent, alternative is to charge a two part tariff with a fixed fee of $3 per customer and then to charge a price of $1 per unit consumed. Note that with this form of price discrimination the monopolist doesn't need to distinguish class A and class B consumers – the monopolist offers the same tariff to all consumers, who simply select the quantity they wish to purchase, revealing *ex post*, the class to which they below.

Example 3:

The presence of a substitute can disrupt the ability of a monopolist to differentiate its prices and, in an extreme case, force it to shut down. Consider again the example above, with one good and two classes of consumers. Suppose, in addition, that the monopolist also produces another good which is a perfect substitute for the existing good, but suppose that price discrimination in this good is impossible – that is, the monopolist can only sell it at a simple, linear price.

As noted above, in this example there is not one unique efficient set of prices – rather there is a range of efficient prices. For example, the following prices were efficient: selling at price $4 to class A consumers and $2 to class B consumers; alternatively, selling at price $7 to class A consumers and $1 to class B consumers. The price of the substitute represents a ceiling on the price at which the goods can be sold to the higher priced customers – *i.e.*, class A customers. The lower the price of the substitute, the smaller the range of efficient prices.

If the price of the substitute good is less than $4 there are no prices at which the monopolist can earn sufficient revenue to cover its fixed costs. Suppose that the price of the substitute good is less than $4. In this case, class A consumers would then switch to consuming this good, whether the price discrimination is in the form of a two part tariff, or simply different prices to different consumer groups. This results in a contribution to fixed cost from class A consumers of less than $200, but class B consumers cannot yield a contribution of more than $300 (that is their total consumers' surplus), so the monopolist is unable to break even.

Principle 7: (Price Caps) *The regulator can reduce its own information requirements and make use of any information held by the regulated firm by delegating the setting of individual prices to the regulated firm, while requiring that the weighted average of the prices set by the regulated firm does not exceed a certain level. If the weights and the level of the cap are set correctly, the regulated firm will use the information* *that it has to set the correct, efficient, level for the regulated prices.*[12] *Setting the weights correctly may be difficult. In addition, if price discrimination is possible on some services but not substitute services, a regulated firm subject to a price cap may choose to withdraw from producing certain services.*

2.2. Principles for regulation of the access prices of a natural monopoly

We are now in a position to apply the principles set out above to the problem of regulating access prices. We will assume that the downstream industry is perfectly competitive. This assumption has the effect of allowing us to treat the downstream industry as though they are "final" customers – with no distinction between sales to rival firms and sales to final consumers. The above principles can then be applied without modification. The problem of access pricing reduces to the problem of determining the efficient prices for a natural monopoly.[13]

Because the access service is not sold directly to final customers, but to downstream firms which transform the product before selling it on to final customers, the demand for the access service is derived from the demand for the final service that is produced using the access service. We will use the phrase "corresponding final service" to refer to the final service produced by a downstream competing firm using the access service. We can then say that the demand for an access service is derived from the demand for the corresponding final service. The phrase "corresponding final customer" will similarly refer to the final customer who purchases the corresponding final service.

A primary issue to keep in mind when applying the principles set out above is that often the corresponding final services (*i.e.*, the services produced by the downstream firms using the access input) will be substitutes for the final services of the monopolist itself. Therefore, it is important to take into account the effects on substitution between final services when calculating the efficient set of prices of the monopolist. In summary:

The one-way access pricing problem is essentially the standard problem of the regulation of the prices of a natural monopoly, where account is taken of the fact that the corresponding final services are often substitutes for services produced by the monopolist itself. When this fact is taken into account, the access pricing problem is simply an application of the principles set out above.

We will proceed, therefore, by simply applying the principles set out above. In the case where there is no need to recover the total costs of the monopolist through its prices, or where any deficit can be recovered efficiently through other means, and when all access and final prices of the regulated firm can be varied freely, the efficient access price is simply equal to the marginal cost of access. This result is derived in the appendix to this chapter.

Principle 1 (again): In the absence of other constraints or objectives, such as the requirement that revenues cover total costs, efficient access prices (and final prices) of the monopolist should be equal to their respective marginal cost.

Although marginal-cost-based access pricing is unknown in some sectors it is not uncommon in others. In particular, it is not uncommon in the transport sector, for example in pricing access to the track infrastructure in the rail sector.

As principle 2 emphasises, if one or more final prices of the regulated firm are set above marginal cost to satisfy other objectives, the final prices of substitute services must also be above marginal cost to efficiently preserve these objectives. This can be achieved by ensuring that the prices of access services are above marginal cost by an appropriate margin or by taxing final services.

If the regulated firm only produces two services, an access service and a final service, and the final price is fixed away from marginal cost, then it is straightforward to show that net revenue can be efficiently maintained at the level that satisfies the objectives by charging an access price that is equal to the marginal cost of access plus another term which is equal to the profit from selling an additional unit of the final product multiplied by the degree of substitution between the monopolist's final service and the competitors' service. See equation 3 in the appendix. Alternatively, if taxes on the monopolist's and competitors' final service are available instruments then access can be priced at marginal cost and taxes on final services can be set by the Ramsey method to meet the price constraint on the monopolist's final service and the net revenue objective. If input proportions are fixed, and competitors cannot bypass the monopoly input, this approach is equivalent to the first with respect to the final service prices, the amount of access purchased, and the amount of revenue raised. In this special case, which corresponds to the assumptions we maintain in most of this report, we use the term "access price" to represent the price charged directly by the monopolist for the monopoly input plus (or minus) any tax (or subsidy) on the final product. However, it is important to remember that when proportions are not fixed, or bypass is an issue, using explicit taxes and subsidies on final services maintains incentives to make efficient use of the monopoly input.

In the special case where the competitors' services are a perfect substitute for the incumbent's services, the access price is simply equal to the monopolist's final price less the marginal cost to the monopolist of turning a unit of access service into a unit of final service. This is one way of stating the well-known Efficient Component Pricing Rule ("ECPR"). The case of perfect substitutes has particular relevance for access pricing for the following reason: If the total quantity of inputs that the monopolist can provide is fixed, then for the purposes of setting access prices it is necessary to treat the product produced

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

by the downstream firms as a perfect substitute for the product produced by the monopolist – whatever the service produced by the downstream firm! The reason is that if the total quantity of inputs is fixed, an increase in the quantity of final goods produced by the competitors always comes at the expense of a reduction in the quantity of final goods produced by the monopoly.[14]

Principle 2 can therefore be viewed as providing a straightforward generalisation of the ECPR to the case where the final service of the monopolist and the final service of the rival are not perfect substitutes. See equation 4 in the appendix.

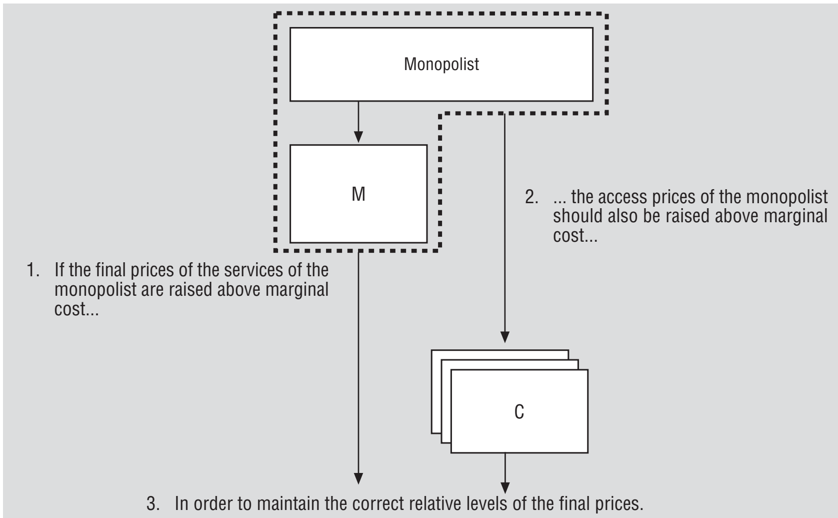
The following principle merely states that economic efficiency requires that if a retail price is above marginal cost (for whatever reason) then the total amount paid for access (which is called here the access price) must also be above marginal cost (in the case of substitutes) to maintain net revenue.

Principle 2 (again): If the final prices of the monopolist are above (or below) marginal cost the final prices of substitute services produced by downstream firms must also be above (or below) marginal cost and therefore the corresponding access prices must also be above (or below) marginal cost. If the final services of the downstream firms are perfect substitutes for the services of the monopoly, the access price must be such as to ensure that the corresponding final price is equal to the final price of the monopolist. The corresponding access price is in this case given by the simple form of the ECPR rule.

It is important to be clear what this principle is stating. A common objection to the use of the ECPR rule is precisely that it preserves the existing (possibly inefficient) structure of the incumbent's retail prices. This is true. Principle 2 simply states that *if the regulator wishes* to preserve the existing (possibly inefficient) structure of the incumbent's final retail prices then the regulator must adjust the access prices along the lines given by the ECPR rule. Of course, if the regulator does not wish to preserve the existing structure of retail prices, it would be wrong to adjust the access prices in this way. Indeed, the regulator may be able to improve efficiency by encouraging the elimination of an inefficient retail price structure precisely by not adjusting the access prices in this way and instead basing the access prices directly on the underlying marginal cost.[15]

Principle 2 applies both when the final prices of the incumbent are either above or below marginal cost. When the final prices of the incumbent are above marginal cost, the access prices corresponding to substitute final services should be above marginal cost (and the access prices corresponding to complementary final services below marginal cost). Conversely when the final prices of the incumbent are below marginal cost, the access prices for corresponding substitute final services should also be below marginal cost.

Figure 2. **Link between access prices and final prices when final prices are different from marginal cost**



Source: OECD.

As mentioned earlier, in most cases, the total revenue of the monopolist must exceed the level that would result from marginal cost pricing. Applying principle 3 above gives the correct mark-up of final and access prices over marginal cost. In particular, access prices should be increased above marginal cost by an amount given by the Ramsey formula. This formula takes into account substitution effects between services and, in particular, between final services and access services. In the extreme case, if corresponding final services are a perfect substitute for the final services of the monopolist then the complex Ramsey formula reduces to the simple ECPR formula mentioned above. This is shown in the appendix.

Principle 3 (again): If marginal cost pricing yields insufficient revenue to the monopolist then both access and final prices must be raised above marginal cost according to the Ramsey formula, or taxes and subsidies must support the differences between the final prices, determined by the Ramsey formula, and the access prices.

In the telecommunications industry the single most common approach involves not a mark-up on the marginal cost of providing access services but a mark-up on the incremental cost of access services. This approach be can justified on the grounds that if any service (or group of services) earns a revenue below its incremental cost then it must be that some other service (or group of services) earns a revenue above the stand-alone cost. If it is feared

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

that pricing above stand-alone cost will induce inefficient entry, the regulator may wish to ensure that the revenue for any service (or group of services) exceeds the incremental cost of that service. This may be the explanation behind the virtually ubiquitous use of long-run incremental cost as a benchmark for setting telecommunications access prices (in the form TELRIC or TSLRIC).

However, in most cases the entrant will not be able to enter even if it can earn the stand-alone costs of the incumbent (for example, the entrant will typically only capture a smaller share of the market and therefore will not be able to benefit from economies of scale to the same extent as the incumbent) so the regulator has some flexibility to allow revenues for groups of services to rise above stand-alone cost. Doing so will, in some cases, increase total welfare. For example, where some services have highly inelastic demand, reducing the regulated price for other services down to marginal cost (not incremental cost) will increase total welfare.

As principle 4 indicates, when it is possible to price discriminate in access or final prices it will often be efficient to do so. In particular, if it is possible for the monopolist to determine the identity of the final customer who purchases the corresponding final service from one of its competitors, and if it is possible to prevent resale between final customers, it is usually efficient for the monopolist to set different access prices for different (classes of) final customers. Similarly, if it is possible for the monopolist to determine the date, time or geographic location of the final service provided by one of its competitors, then it is usually efficient for the monopolist to set different access prices according to these characteristics. Finally, if it is possible for the monopolist to measure the amount of final service consumed by the customers of its competitors, it is usually efficient to set the access price accordingly. In particular, it may be possible to use a two-part tariff for the access price.

Note that such price discrimination at the access level requires knowledge by the regulated firm of the identity of the final customer and *not* just the identity of the firm purchasing the access input. In other words, the regulated firm must be able to trace to whom (or when or where) the corresponding final service is sold. This might not always be possible. We discuss further below the implications of the inability to price discriminate effectively at the access level.

Is there a danger that the use of two-part prices at the access level will restrict competition downstream? After all, don't two-part access prices introduce a form of increasing returns to scale downstream limiting the number of firms that can survive – perhaps even creating a downstream monopoly? Indeed, in Finland, in response to liberalisation of the local loop, the incumbent operators introduced substantial discounts for volumes (a form of two-part pricing) that only their own affiliates could obtain, effectively excluding rivals.[16]

The use of two-part access prices will not restrict competition downstream if the fixed component of the two-part tariff depends proportionally on the number of customers served by the downstream firm, and not on the identity of the downstream firm. As long as the fixed charge is proportional to the number of final customers, the tariff confers no advantage on firms serving more customers. Of course, an access charge structured in this way requires knowledge of the number of customers served by a downstream firm and (at least) their total consumption. As mentioned earlier, this information might not always be available.[17]

Principle 4 (again): When price discrimination in access pricing is feasible it is usually efficient to price discriminate. In other words, it is usually efficient to set different access prices according to the identity of the corresponding final customer or the date, time or geographic location of the sale to the corresponding final customer, or the number of units purchased by the corresponding final customer. To do this requires information about the identity of the corresponding final customer even though the access good is not sold directly to the corresponding final customer but to a downstream firm.

As noted earlier, price discrimination is equivalent to extending the range of services sold by the monopolist. The access price for each of those new services should be set according to the principles enunciated earlier if the goal is to efficiently raise revenues in excess of those collected by pricing at marginal cost. In the case of perfect substitutes, the access price should be such as to ensure that the contribution to any such revenue requirement is preserved whether or not the monopolist sells access rather than its own final service. Since the contribution to the revenue requirement depends on the price-cost margin for the service, the access price will, in turn, depend on the revenue earned from the service before and after granting access and the costs incurred by the monopolist before and after granting access. This principle applies to both the fixed and the access component of a two-part final price. See the first few examples in Box 2.

Assume now that direct taxes on final services cannot be discriminatory. Suppose that, while it is possible for the monopolist to price discriminate in its own sales to final customers, it is *not* possible to determine the identity of the corresponding final consumer of the regulated firm's competitors, so that it cannot price discriminate as efficiently in its sales of the access service. Suppose also that the monopolist's final service and the corresponding final service of its competitors are substitutes. In this case the inability to price discriminate at the access level interferes with the ability of the monopolist to price discriminate at the final level (principle 5).

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

In particular, the inability to match the price discrimination at the final level to price discrimination at the access level leads to a conflict between objectives. Raising the desired revenue efficiently requires price discrimination, but this price discrimination is undermined by the inability of the monopolist to price-discriminate at the access level or of the regulator to levy the correct taxes at the final service level. If the desired revenue is to be collected efficiently, it is essential to raise the price at the access level (to at least the level of the highest price of the corresponding final service) or to withdraw the access service from sale entirely. Obviously both of these possibilities restrict the scope for competition from new entrants.

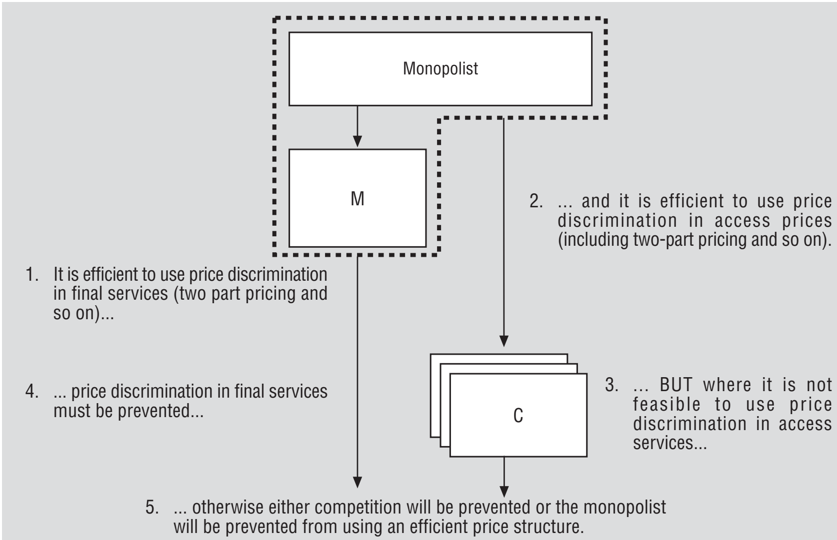
On the other hand, if the access price is to be maintained at a level that allows a reasonable scope for entry then the price discrimination by the monopolist in its final services will be undermined. The monopolist will not be able to raise the price of its final services above the cost of the corresponding final services of the competitors, even when it is efficient to do so. Thus, either competition or the ability to efficiently raise revenue is undermined.[18] There are several concrete examples of this in Box 2.

Principle 5 (again): If the regulated firm cannot, for the purposes of setting the access price, obtain information about the identity or number of units purchased of the corresponding final customer or the date, time or location of the corresponding final transaction, or if the regulated firm cannot prevent resale amongst rival downstream firms, fully efficient price discrimination at the access level is not feasible. In this case there is a conflict of objectives. Efficient final price discrimination requires raising the access price, limiting or preventing downstream competition. On the other hand, fixing a "reasonable" access price may undermine an efficient final price structure.

One common situation that is worth mentioning is the case where the monopolist cannot identify the corresponding final customer of its competitors but it can nevertheless identify the *competitor* to whom it sells a particular access service and it can prevent that competitor from on-selling that service to other firms. In this case, the monopolist can still use some forms of price discrimination – in particular, the monopolist can use a form of two-part pricing where the fixed part depends only on the identity of the downstream firm (in the discussion above the fixed part depends on the identity of the corresponding *final* customer).

Ordinary two-part pricing introduces increasing returns to scale in the downstream competitive industry and therefore runs the risk of turning this industry into an oligopolistic or even monopolistic industry. Ordinary two-part pricing at the access level is therefore usually undesirable. Capacity-based access pricing, on the other hand, doesn't have this drawback. Capacity-based access pricing involves the use of a two-part tariff. The marginal price for

Figure 3. **Price discrimination in access prices and final prices**



Source: OECD.

usage is set equal to the (typically low) marginal cost for usage as long as the competitive firm is purchasing less than a certain level of capacity. The fixed component of the two-part tariff is made proportional to the maximum quantity that the downstream firm can purchase at the low marginal cost. As noted earlier, this approach confers no advantage on downstream firms serving a larger number of customers.

Capacity-based pricing has the effect of dividing up the natural monopoly into a number of smaller parts, which may then compete with each other. Capacity-based pricing may be combined with long-term contracts with the regulated firm. Such longer-term contracts require the downstream competitive firms to bear some of the risk that the regulated firm incurs in investing in the additional capacity necessary to serve the downstream industry. Long-term contracts at capacity-based prices are, in fact, a form of club or joint ownership of the essential facility.

One of the principal advantages of capacity-based pricing is that it allows the downstream competitive firms to break away from the tariff structure produced by regulation. This may be useful if either regulators are not able to respond efficiently to real time variations in market conditions or regulators do not appropriately value efficient prices. Capacity based pricing achieves this by allowing the downstream firms to, effectively, purchase a share of the essential facility, "scaled down" to match the requirements of the downstream firm, but with the same cost drivers as are faced by the incumbent operator.

Capacity based pricing is common in the US natural gas industry (where long-term contracts for pipeline capacity are common) and in allocating take-off and landing rights to airports. In the telecommunications industry, where marginal cost is close to zero for a network operating below capacity and for which the fixed costs of establishing a network are closely related to the capacity of that network, capacity-based pricing has many natural advantages. Capacity-based charging is, of course, very common in pricing wholesale connections to the Internet. The European Commission writes:

> "A fixed network has to be dimensioned according to the quality of service required during the period of peak traffic load which it will be required to handle. If a network has to handle peak hour traffic from other interconnected networks, additional capacity will be necessary to maintain the desired quality of service. An analysis of the capacity costs required to provide that service quality (given demand, or subsequently, actual interconnect traffic figures) will enable costs to be apportioned among interconnecting parties.
>
> In an ideal situation, where an industry comprised established market players with relatively stable market shares, capacity based charging would be the most efficient interconnect pricing rule."[19]

Since volumes of telecommunications traffic are not uniform over the course of a day or week, the maximum capacity of the network may only be reached during one hour per week or per month. If different firms have a different time profile of traffic, the charge for a use of capacity should vary according to the time of day and day of the week, to reflect the different degree of network congestion at different times. It may be possible to establish markets in capacity rights at different times. This creates incentives for network operators to make use of the network at off-peak times and ensures that the capacity costs are primarily borne by those networks that require capacity during the "busy hour" period.

In practice, the number of markets required could be very large indeed – one for each hour of the day at each point of interconnection. The costs of establishing such markets could be prohibitive. An alternative, less theoretically pure approach is "retrospective charging" where each network is charged for its share of the total traffic during the busiest hour over a period of, say, a month. In the mid 1990s Mercury (a new entrant in the UK) proposed to hold a trial of retrospective capacity-based charging in its interconnection agreements with cable operators. As the next chapter of this report discusses, capacity-based charging is becoming increasingly common in the telecommunications, particular in the provision of call origination services to Internet service providers.

As principle 6 indicates, it is not unusual for the above principles to yield an access price for a particular service or group of services sufficiently large to induce firms to bypass the monopolist and to produce the service themselves. As long as the industry is a "natural monopoly", such entry by other firms is likely to involve higher costs and should be deterred. Unfortunately, in the absence of some additional instrument, such entry cannot be deterred except by lowering final and access prices on those goods threatened by inefficient entry, which may conflict with other objectives such as allocative efficiency. Furthermore, principle 5 indicates that downstream competition could prevent the monopolist from pricing access to support price discrimination in final services even where the competitors continue to buy access from the monopolist.

As discussed earlier, these conflicts might be resolved through a tax on final services. For example, a tax could be imposed on providers of certain final services (whether they purchase the corresponding inputs from the monopolist or not). Efficient tax levels would be set according to the final consumers known characteristics (*e.g.*, consumption level, date, time, place, and class) as is typically done in promulgating rates under regulation. The corresponding access prices would then be moderated to a level that does not induce inefficient entry (*e.g.*, such that revenue is below standalone cost). Explicit taxes of this kind could be implemented in the form of a universal-service type fund. Such taxes and access charges can be "competitively neutral" with respect to monopolist and entrants alike so as not to artificially distort customer purchases.

Principle 6 (again); The above principles may lead to access prices which induce inefficient entry. In the absence of additional instruments, control of this entry may conflict with other objectives such as allocative efficiency. Where another instrument, such as a tax/subsidy or "universal service" mechanism is available, it is possible to achieve both of these objectives without conflict.

Finally, as noted above, the problem of setting the access and final prices of the monopolist can, in some circumstances, be devolved to the monopolist itself, by allowing the monopolist to choose its own prices subject to a cap on the weighted average of prices. This principle applies to access prices – only it is essential that both access and final prices are included within the overall price cap. Laffont and Tirole refer to this as a "global price cap". A monopolist subject to a global price cap (with the correct weights on each service) has, in principle, the correct incentives to choose an efficient, welfare-maximising structure for access and final prices.[20]

One benefit of the global price cap especially deserves to be noted. The incentive on the monopolist to prevent or deny access depends on, amongst other things, the relative "tightness" of the regulation of final prices relative to access prices. If access prices are relatively "tightly" regulated (*i.e.*, held close to marginal cost) while final prices are loosely regulated or unregulated, the monopolist has an incentive to deny access in order to limit downstream competition, to earn greater rents in the downstream market. The opposite is true when access prices are loosely regulated and final prices are tightly regulated. One of the benefits of the global price cap is that, by treating access and final prices symmetrically, there is no incentive for the monopolist to prefer to sell at the access level or at the final level. In principle, this eliminates the incentive to restrict or deny access.[21] Put another way, under the global price cap, the monopolist can actually make more money by selling access to a more efficient downstream rival than by producing the downstream product itself.

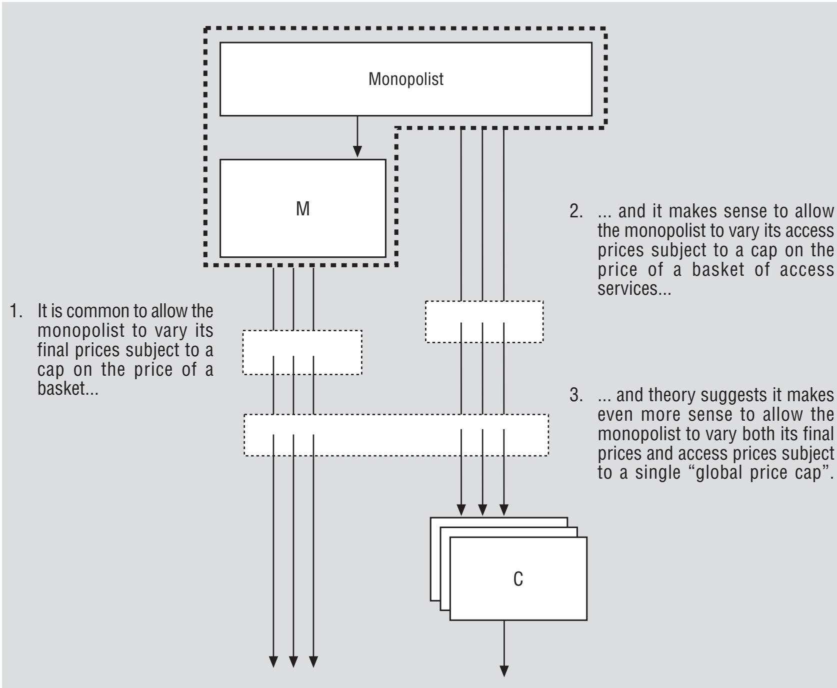
However, as noted earlier, when price discrimination is possible at the final level but not at the access level, the monopolist has an incentive to raise the access price, not only to limit downstream competition, but also to preserve the ability to discriminate downstream. This incentive remains under a global price cap. Under a global price cap, if price discrimination is not possible at the access level, the monopolist will prefer to set a higher access price, not to reduce competition but to preserve its ability to price discriminate to final customers.

In addition, the motive to exclude may still exist in a longer-run, strategic sense. For example, entry might produce cost information that would cause the regulator to tighten the price cap. Entry could threaten the monopolist's grip on the underlying monopoly product, or reduce a network monopolist's ability to harvest benefits from network externalities. Entry could reduce the profitability of shading quality under a price cap on final goods. These caveats about the global price cap are theoretical in the sense that the global price cap is a relatively new idea that has not been implemented in practice. However, if the motive to exclude or predate exists, the global price cap, by allowing the monopolist to offset price reductions at the final service level with access price increases, might facilitate predation.

A compromise approach is to allow the regulated firm some flexibility to choose its prices, but with separate price caps on final services and access services. Although a full global price cap has not (yet) been tried, separate caps on final prices (and sometimes access prices) are found in a few countries (see the next chapter of this report).

Two potential impacts devolving pricing flexibility to the monopolist also warrant comment. A well-known benefit of such flexibility is that any superior

Figure 4. **Price caps on access services and final services**



Source: OECD.

knowledge of demand or cost conditions on the part of the monopolist will be incorporated into prices. A potential, drawback is that entrant's incentives to innovate can be reduced if the monopolist uses its access pricing flexibility to expropriate entrant's innovations.

Principle 7 (again): A regulator can usually improve efficiency by allowing the regulated firm to vary its access prices subject to a cap on a weighted-average of a basket of prices. In theory, although this has yet to be tried in practice, the incentive on the monopolist to restrict competition can be reduced through the use of a global price cap covering both access and final services.

Imperfect competition downstream

Throughout this section of this report we have assumed that the downstream sector is perfectly competitive. In practice, in many utility industries, although the downstream sector is somewhat competitive, the resulting level of competition seldom approaches the perfect competition ideal. What effect does the lack of perfect competition downstream have on the theory set out above?

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

This discussion has highlighted the importance of maintaining the correct relative level of the final prices of the competitors and the incumbent. This principle continues to apply even when there is imperfect competition downstream. A primary consequence of imperfect competition downstream is that the final price faced by end-users of the competitors is raised above the marginal cost of providing the service of the competitors – which includes the cost of access. This mark-up above marginal cost is inefficient for the usual reasons (it artificially reduces consumption and distorts consumption in favour of other products). Lowering the access charges relative to the level given by the previous principles can eliminate the price-marginal cost mark-up on the final services. It may even be necessary to lower the access charges below the marginal cost of access.

Examples in the telecommunications sector

In order to make these principles more concrete, this section illustrates how these principles would be applied in the telecommunications industry through the use of several worked examples. Most of these examples will focus on local loop unbundling ("LLU"). LLU arises when the incumbent leases the local loop at wholesale prices to a rival to enable the rival to provide a range of telecommunications services.

From the previous principles we know that one key factor is whether the downstream firm will use the unbundled local loop to provide a service which is a substitute for, a complement to, or independent of the incumbent firm's existing services. In the case where the new service of the entrant does not compete with the service of the incumbent,[22] the principles set out here show us that the existing retail tariffs of the incumbent for its services are irrelevant. The access price for the new service should be based on its marginal cost, with (if necessary) a mark-up to provide a contribution towards fixed costs based on its own elasticity of demand. If this access price is so high as to induce stand-alone entry into local loop services, it may be necessary to either lower the access price or to introduce some other instrument, such as a tax on the services provided with local loops.

In the unlikely case that the new service is a complement for a service of the incumbent, the access price for the new service should take into account the retail prices for complementary services. In particular, the access price should be lowered relative to the price given in the previous paragraph in order to stimulate demand for the service and its complements.

The most likely case is where the new service of the entrant is a substitute for the service of the incumbent.[23] In this case, the above principles highlight the importance of taking into account the relationship between the access prices and the retail prices of the incumbent. In the case of perfect substitutes (which may be the most likely case) the relationship between the access prices and the retail prices of the incumbent should be given by the simple form of the ECPR. Specifically, the access price for unbundled local loop should be equal to the marginal cost of providing the local loop plus an appropriate contribution to fixed costs earned by the incumbent on its use of the local loop. In the case of perfect substitutes the correct relationship between access prices and retail prices can be achieved either by taking the retail prices as given and setting the access prices accordingly, or by setting the access prices and allowing competition to drive the retail prices to the required level.

In addition, if the structure of retail prices is to be preserved, any price discrimination that is present in the retail prices should be reflected in the access price for unbundled local loops. This implies that:

1. Where the price for final services which use the local loop has a two-part tariff (with the usage part above marginal cost) the access price for unbundled local loop should also use a two-part tariff. For example, if, say, the retail price of call services, involves both a monthly charge and a per call charge, then the price for unbundled local loop should also have this two-part structure.
2. Where the price for final services that use the local loop distinguishes between different classes of customers, the price for unbundled local loop should also distinguish between classes of customers. For example, if business customers pay a higher monthly fixed charge than residential customers, this should be reflected in different prices for the unbundled local loop.

Finally, where stand-alone provision of local loop services is feasible, the access prices for a set of local loop-based services should not yield revenue sufficient to induce inefficient entry. If retail prices do not vary with geographic location (even though costs vary) it is quite likely that the prices for local loop services in densely populated areas will exceed the stand alone cost of providing those services. In this case another instrument is likely to be required to prevent inefficient entry. For example, the access price could be lowered to a level that does not induce entry and providers of the corresponding retail service required to pay a tax or contribution to a common fund, as described above.

Similar analysis can be applied to the problem of mandatory roaming. Economies of scale in mobile services are likely to be more important in remote (low density) areas. If mobile users are prepared to pay more for mobile operators with greater geographic coverage, mobile operators can cross-subsidise service in remote areas from some of the revenue from service provided in high density areas. But, it is occasionally argued that the level of

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

mobile competition in high-density areas could be increased if new entrant mobile operators could obtain access to the networks of existing operators in remote areas. The principles developed here can shed some light on how that access should be priced. If the mobile service of the rival mobile operators is a good substitute for the mobile services of the access provider (*i.e.*, the mobile network with full geographic coverage) then the sale of roaming services in remote areas may affect demand for the access provider's network in all areas. This should be taken into account when setting the access price in remote areas. Specifically, the access price should reflect the lost profits from loss of business in densely populated areas. See example 5 in Box 3.

3. Two-way access pricing: interconnection of many competing networks with one central network

We turn now to the problem of two-way access pricing. This situation arises when two (or more) companies must purchase essential inputs from each other. This arises commonly in the telecommunications industry. The provision of ubiquitous communication services often requires access to another company's network in order to be able to provide communications services to a person attached to that other network. For example, to provide ubiquitous services, mobile network operators require connection to fixed network operators and *vice versa* . Fixed network operators in one country require connection to fixed network operators in another country and *vice versa*.

Two-way access issues also arise in other industries. For example, in the rail sector, two networks can offer an expanded range of direct connections between origination and destination cities if they allow trains to operate over each other's tracks. In the postal sector, the provision of ubiquitous mail service typically requires that mail collected by one company be delivered by another and *vice versa*.

Models of two-way access problems quickly become very complex, involving complex interactions of assumptions regarding the nature of competition between different networks that are offering different services, with differing assumptions about the constraints imposed by regulation. As a result, models of two-way access in the economic literature focus on specific special cases. Although some hints are possible, it is not always possible to generalise the implications of specific cases into general principles as we have done in the first part of this paper.

In this part of this chapter we will deal with a specific form of two-way access issues – where competing downstream firms need to purchase inputs from the monopolist, and *vice versa* , but the downstream firms do not need to purchase inputs from each other. In the final part of the chapter, we will look at models where competing firms purchase inputs from each other.

Box 3. **Price discrimination in access prices and final prices**

The following examples are intended to highlight the problems that might arise when price discrimination occurs in final prices but not in access prices.

Example 1:

Suppose that a telecommunications company charges a monthly fee for telephone service that depends on the class of customer. Suppose that there are two classes of customers – "residential" customers who pay $80 per month and "business" customers who pay $120 per month. The cost of providing local loops to each of these customers is $100. There are equal numbers in these two classes. These prices are chosen so that the telecommunications company just breaks even.

If we assume that the incumbent still incurs the cost of the local loop after granting access, the principle of covering the contribution to fixed costs provides that the appropriate access charge should discriminate between the two classes of customers, with a charge of $80 for residential customers and $120 for business customers.

Suppose that the regulator decides to unbundle the local loop at an undifferentiated price of $100, without discriminating between business and residential customers. With these prices, the new entrants will clearly target business customers, driving down the monthly rental for business customers. The incumbent will be left with a higher proportion of residential customers. If the incumbent must provide service and is not allowed to raise its prices to residential customers, granting access threatens the profitability of the incumbent company.

The scope for competition for both classes of customers could be restored either by differentiating the access charge or by introducing some form of fund (such as a universal service subsidy scheme) which achieved the same end. For example, firms might be required to pay an additional $20 into a fund for each business customer they serve, and would be allowed to receive $20 from the fund for each residential customer they serve. In this way, the "effective access price" (*i.e.*, the access price plus any universal service tax or subsidy) fully reflects the differentiation in retail charges, as required.

Of course it may be undesirable to preserve the existing rate structure. Placing both residential and business retail lines under the same price cap and allowing residential rates to rise as competition or profit maximization by the incumbent drove business rates down does not preserve the existing retail rate structure but still permits competition without being confiscatory. This alternative would permit an increase in retail pricing efficiency.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Box 3. **Price discrimination in access prices and final prices** *(cont.)*

Example 2:

Suppose a local loop in a certain region costs $120 (including any fixed costs that need to be covered). Suppose that the incumbent telecommunications company charges a two part tariff for telecommunications services – the fixed subscriber charge of the incumbent is $100 and the incumbent's usage charge per call is $1. Suppose, finally that there are two types of consumers, type A, which comprise 20% of the population which are high users, consuming 60 units of usage and type B (80% of the population) which consume 10 units of usage. Assume that the marginal cost of providing usage is zero.

With this structure of prices and costs the incumbent breaks even – the incumbent recovers 100 + 60 × 1 = $160 from the high usage customers and 100 + 10 × 1 = $110 from the low usage customers, which gives an average revenue per line of 20% × 160 + 80% × 110 = $120 – sufficient to cover the local loop cost of $120.

The principles regarding recovering the fixed costs tells us that, if the retail rate structure is to be preserved, the appropriate access charge in this context is a two part charge with the fixed component equal to $100 and the usage component $1.

Suppose that the regulator does not adopt this structure of access prices. Specifically, assume that the unbundled local loop price is fixed at the undifferentiated flat fee of $120. With this local loop price, an entrant who targeted a type A customer would receive a profit of 160 – 120 = $40, or drive price for type A customers down to $120 if competition were robust. On the other hand, an entrant who targeted a type B customer would receive a profit of 110 – 120 = $10. Clearly, the entrants will be limited to the market for high usage customers. The incumbent is left carrying only the low usage customers for whom the average contribution of $110 is less than the incumbent's average cost of $120.

As before, a suitably designed fund (such as a universal service scheme) might be able to restore the scope for competition for both classes of customers. For example, the fund might be set up in the following way: the fund could receive $1 for each unit of usage and pay out a flat fee of $20 to each customer. Firms would then expect to make a net payment to the fund of $40 for the high usage customers and to receive a net payment from the fund of $10 for the low usage customers. Again, the effect of the fund is to ensure that the "effective access price" fully reflects the differentiation in retail prices.

Box 3. **Price discrimination in access prices and final prices** *(cont.)*

Example 3:

Suppose that there are two types of telecommunications consumers – high users who always stay on line for 12 minutes and are prepared to pay up to $5 for the call, and low users who only stay on line for 2 minutes and are not prepared to pay more than $1. A telecommunications incumbent, recognises this and offers a menu of prices – specifically, the company charges 50 cents per minute for calls, but places a cap of $5 on the price of a call. Suppose that the cost of producing a call is 42.86 cents per minute and that there are equal numbers of high and low users. It is easy to check that the average revenue per consumer is $3 and the average time on line is 7 minutes, so the incumbent just breaks even ($3/7 = 42.86 cents).

The principles set out above suggest that the access charge should be a two part charge with a fee of 50 cents per minute up to ten minutes and then zero thereafter. With this access charge, the differentiation in retail prices is fully reflected in access prices.

Suppose that the regulator sets a simple linear access price of 42.86 cents per minute. When a new entrant, charging the same prices as the incumbent, wins a low user, its profit is 2 × 50 cents – 2 × 42.86 cents = 14.28 cents. On the other hand, when the new entrant wins a high user, its profit is $5 – 12 × 42.86 = 14.32 cents. In this case the entrant will target only low users, limiting the scope for competition and threatening the profitability of the incumbent. As before, the scope for competition can be restored without changing the retail rate structure either by adjusting the access or retail prices (so that the differentiation in one is reflected in the other) or by a suitably designed fund (in this case the fund would tax the first ten minutes of usage and would subsidise longer calls).

Example 4:

Suppose that local loop in a certain region costs $100 and suppose that the incumbent firm offers two calling plans – a low user plan which has no fixed fee and a usage fee of $6 per call, and a high user plan with a fixed fee of $100 and a usage fee of $1 per call. Suppose that the marginal cost per call is $1. Suppose that the regulator wishes to introduce competition for the call component – if the rival handles a call it also incurs the cost of $1. The principles set out above suggest that an access price that efficiently preserves the existing rate structure is $5 per call for a low user and $0 for a high user.

Suppose that the regulator does not differentiate the charges for high and low users. In particular, suppose the regulator sets the same price for everyone.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Box 3. **Price discrimination in access prices and final prices** *(cont.)*

Then clearly there will be an incentive for all users to select the low user plan, avoid the fixed fee and then to choose the rivals to carry the calls. The incumbent recovers no fixed cost at all and is forced to go out of business. In order to prevent this outcome, the regulator may be led to prohibit competition, rather than drop the low user scheme. This point is emphasised by Laffont and Tirole (2000), page 111:

"A hotly debated issue is whether the entrants should pay a differentiated access charge when obtaining access to the customers subscribing to the incumbent's low and high user schemes. Standard non discrimination rules as embodied in various directives and laws as well as the precept of cost based access pricing suggest that such discrimination is unwarranted… Regulators in the UK and France … have allowed the incumbent to foreclose the entrants' access to those customers electing the low user scheme. This regulation … has naturally infuriated the entrants, who complain that they are barred from competing for a substantial number of consumers. Though the entrants definitely have a point, the regulators and the incumbents in our opinion can lay out a very sound case in favour of discrimination. … Uniform access pricing deprives the incumbent of its ability to offer an efficient menu of tariffs tailored to the needs of its clientele."[1]

The underlying assumption above is that the original rate structure must be preserved. Instead, it may be appropriate to relax this constraint and collect a $100 per loop fee for each customer served, regardless of provider, with no usage fee for access. These fees correspond to the underlying cost structure. If retail prices are permitted to adjust freely under robust competition, or flexibly under a price cap, retail prices would reflect the underlying cost structure. In the above example this is likely to be the more efficient outcome because no scale or scope economies requiring prices to be above marginal cost have been designed into the scenario.[2]

Example 5: Access charges for mobile roaming

Suppose that an incumbent mobile network has 1 000 customers. A typical mobile customer makes 50 calls in densely populated areas per annum and 10 calls in sparsely populated areas. The marginal costs of a mobile call are zero and the fixed costs of providing mobile service are $130 per annum per customer in densely populated areas and a flat cost of $50 000 in sparsely populated areas. (This difference in the cost structure is essential to the problem – to make the densely populated areas potentially competitive, while the sparsely populated areas remain a natural monopoly.)

The incumbent network charges no subscription charge and a per call charge of $3 per call. (We will assume that discrimination between the price of

Box 3. **Price discrimination in access prices and final prices** *(cont.)*

calls in remote areas and in densely populated areas is not possible). At this price the incumbent breaks even, since 1 000 (60 × 3 – 130) – 50 000 = 0. An entrant invests in a network in the densely populated areas and seeks roaming in the sparsely populated areas. How much should be paid for access to the incumbent's network in remote areas assuming the $3 per call contribution to the incumbent is to be preserved?

Example 5: Access charges for mobile roaming

If the entrant's network is not a substitute for the incumbent's network, access should be granted with a marginal price equal to marginal cost (which is zero in this case).[3] If the entrant's network is a perfect substitute for the incumbent's then granting access involves the loss of $50 per customer. This loss should be reflected in the access price, either through a linear price of $5 per call in remote areas, or through a fixed fee of $50 per customer. (Since mobile companies do not differentiate their prices between remote and densely populated areas the mechanism by which this access price is charged is irrelevant.)

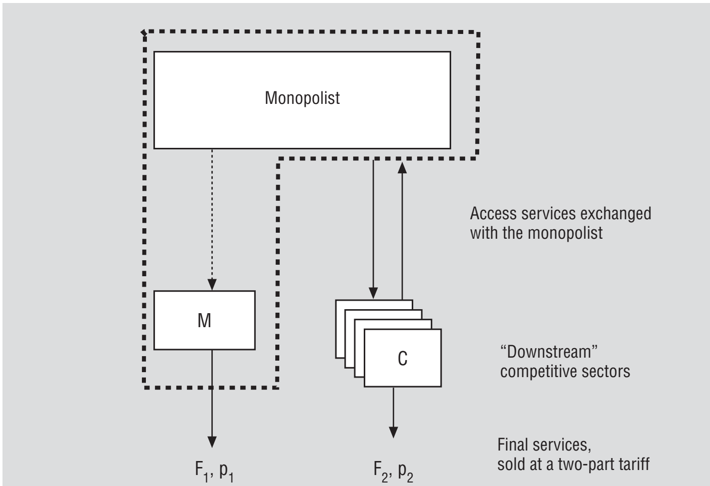
As an aside, note that if competition was expected to lead to an increased number of mobile customers, the fixed fee per customer should be lower than $50, and set at a level to ensure that the incumbent is able to earn $50 000 overall.

Note that the access price must be above the incumbent's retail price in this case – if the access price were set at $3, the incumbent would receive $30 in revenue (instead of $180), and would save $130 in costs, so the contribution to fixed costs would be down $20 per customer lost. With this price the incumbent would be unable to break even.

1. Laffont and Tirole (2000), page 112.
2. In a network industry, there may still be an efficiency case for some price discrimination even without cost economies. Some customers on the network may not value their service as much as is costs to produce it. If parents, relatives and others that would like to talk to them would not fill this value-cost gap by voluntarily subsidizing their phone service, these customers would drop off the network under cost-based pricing. Deviations from marginal cost pricing that keep more customers on the network could be efficiency enhancing if the surplus gained from calls with these customers is greater than that lost as a result of some higher-than-marginal-cost prices.
3. Assuming some demand elasticity for both the incumbent's and the entrant's product, and if the $3 per call contribution was not fixed, the same total contribution to fixed costs could be achieved more efficiently by reducing the $3 per call price on the incumbents calls and taxing the new demands for the entrant's products based on their own elasticity of demand.

We will begin, then, with a model of two-way access that is a straightforward extension of the model used in the previous section on one-way access. As in the previous section of the paper we will imagine two types of firms – a monopolist, which is assumed to operate a "central" network, and

Figure 5. **Market structure of a two-way access problem with competing downstream firms exchanging inputs with a central monopolist**



Source: OECD.

a large number of other competing firms. As before the competing firms purchase essential inputs from the monopolist and combine these inputs with other inputs in order to provide final services to final customers.

The new feature of this model is that now we will assume that the central network, in order to provide a service of its own, must also purchase an input from each of the downstream firms. This assumption introduces the "two-way" element to the model. The competing firms do not purchase inputs from each other. This market structure is illustrated in the following figure.

Although this model is stylised, it can be applied to the interconnection of mobile operators with a common fixed network. The mobile networks need to purchase the service of "call termination" from the fixed network in order to provide mobile services to their final customers. Similarly, the fixed network needs to purchase call termination from the mobile networks in order to provide a ubiquitous telephone service to its fixed customers. It is assumed that all mobile calls terminate on the fixed network (otherwise mobile companies would have to purchase inputs from each other). This assumption, although unrealistic is useful for pedagogical reasons as it allows us to focus first on a subset of the features of two-way network interconnection.[24] We will call this application the "fixed-mobile" application of the model.

In this model, as in the previous model of one-way access, the prices charged by the central monopoly network (including the price for the input sold to the competing firms) are determined by regulation and the prices of the competing firms are determined by competition. But, what about the price at which the competing networks sell their inputs to the central network? Can this price be left unregulated, to be "determined by competition"?

As we will see, under certain circumstances it is not possible to allow the downstream firms to set their own price at which they sell inputs to the monopolist – under these circumstances each of these downstream firms has an incentive to set this price at the monopoly level.

This problem is known as the "terminating network monopoly" problem. In order for a terminating network monopoly problem to arise we need the following conditions:

1. First, the competitive firms must have a monopoly over the essential inputs required by the monopolist (*i.e.*, each competitive firm must have a monopoly over termination of calls to its own customers). In the case of fixed-to-mobile calls, the most common situation is for a final customer to be connected to only a single (mobile) network.
2. Second, the customers of the competing firms must not care about the price of inputs provided by the competing firms to the central firm (in other words, mobile customers must not care about the price of calls made to them) either because they gain no utility from these calls or they do not care about the utility of the callers.[25] If this were not the case, since the number of calls received by a mobile customer will depend in part on the fixed-to-mobile retail price, other things equal, if mobile customers gain utility from calls made to them they will select a mobile network with a lower fixed-to-mobile retail price.

   This assumption effectively rules out the presence of "closed user groups". A closed user group exists when a fixed-network customer makes a number of calls to a mobile customer and pays for the mobile charges of that customer.[26] In this case the fixed network customer pays both fixed-to-mobile and mobile-to-fixed charges and therefore cares about them both.

   As we will see later, the terminating network monopoly problem is more severe when the fixed-to-mobile price does not depend directly on the individual mobile termination charge or if the calling party does not know the identity of the terminating network. Even if mobile customers care about the price of calls made to them, they may still have no incentive to select a mobile network with a low termination charge if the fixed-to-mobile retail price does not depend directly on the mobile termination charge or if the calling party does not know the identity of the terminating network.
3. Third, the customers of the competing firms do not pay the cost of the inputs provided by the competing firms to the central firms (*i.e.*, mobile customers do not pay for mobile call termination – in other words, the

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

calling party pays principle applies). If mobile customers paid for mobile call termination then they would clearly have a strong incentive to select a mobile network with lower mobile termination charges.

4. Fourth, we will assume that it is not possible to impose reciprocity – that the prices for the inputs sold in each direction are the same (*i.e.*, the price for mobile termination is the same as fixed termination). In this circumstance, as discussed in the next part of this chapter, there may be a countervailing incentive on the termination charge that ensures that this charge is kept low.

Given these assumptions, consider the effect of an increase in the price at which the downstream competing firms sell inputs to the central network. (*i.e.*, an increase in the price of mobile termination). An increase in this price may or may not be passed on to consumers of the fixed network in the form of an increase in the regulated fixed-to-mobile retail price. Assume for a moment that it is passed on – so that an increase in the mobile termination charge results in an increase in the fixed-to-mobile price.[27] In this case an increase in the mobile termination charge will reduce the quantity of fixed-to-mobile calls. But this has no effect on the demand for the services of mobile networks. Under the assumptions above, mobile users do not derive utility from fixed-to-mobile calls, nor do they directly pay for fixed-to-mobile calls, nor do they care about the welfare of their callers. In other words, from the perspective of the mobile networks, increasing the price for mobile termination affects the profitability of the mobile termination service but not the profitability of any other services provided by the mobile firms. As a result, mobile operators, who seek to maximise their profits, will choose to set the mobile termination charge at a level that maximises the profitability of the mobile termination service – *i.e.*, at the "monopoly" level.

The "monopoly" termination charge chosen by any one mobile company will depend on the sensitivity of the retail fixed-to-mobile price to changes in its mobile termination charge. If the fixed-to-mobile retail price is insensitive to the mobile termination charge of any one mobile company, the profit maximising price chosen by that company may be very high indeed.[28] This might happen if, for example, the fixed-to-mobile price were regulated so as to be equal to the average of the termination charges of the various mobile companies. If all mobile companies are small, the fixed-to-mobile price is insensitive to changes in any one termination charge.

Note that this result does not rely on the mobile network having a dominant position in the mobile market. Every network, no matter how small, has a monopoly over calls terminating on its own network and, under the assumptions above, has an incentive to exploit that monopoly.

These results can be expressed as a principle:

In the two-way interconnection of networks, if the customers of each network do not receive utility from (or pay for) inputs sold to other networks and if the customers of each network do not care about the welfare of customers of other networks then, if each network chooses the price at which it sells its own inputs unilaterally, then each network will choose a price which maximises its own profits from selling inputs. This price is inefficiently high from an overall welfare perspective.

Since it is not possible to delegate the setting of the mobile termination charge to the individual mobile networks, it is essential to regulate this access price. At what level should this price be fixed?

The mobile termination price in our model potentially affects two sets of prices:

- First, it affects the price of fixed-to-mobile calls directly, since the mobile termination charge is a key element of the marginal cost of providing fixed-to-mobile calls. Raising the mobile termination charge may therefore affect allocative efficiency – by affecting the price of fixed-to-mobile calls relative to their marginal cost and by affecting the price of fixed-to-mobile calls relative to any other substitute services offered by the fixed network.
- The mobile termination price may also affect the prices of mobile network indirectly – increasing the termination charge increases the profits of the mobile network from termination. Competition between mobile networks forces mobile operators to use these profits to lower their other charges. Lower fixed and usage charges, in turn, may increase the penetration of the mobile networks. An increased penetration of the mobile network, in turn, enhances demand for fixed-to-mobile calls.

These effects are discussed further in Box 4.

To summarise, increasing the termination charge will increase mobile penetration, but at the cost of increasing the fixed-to-mobile call price relative to the true marginal cost. On the other hand, lowering the termination charge will lower the fixed-to-mobile call price, but at the expense of lower mobile penetration. This is summarised in the following diagram.

In essence, the basic problem is that there are two objectives being pursued with a single instrument. The first objective is the right balance between the penetration of the mobile network, on the one hand, and the mark-ups above marginal cost in the fixed network, on the other. The second objective is the right level of the fixed-to-mobile call price relative to marginal cost and relative to the prices of any other substitute products of the fixed network.

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Box 4. **A closer look at fixed to mobile interconnection**

Suppose that we have a number of competing mobile networks, which compete over a two part tariff – that is, they compete using a fixed fee and a per minute fee. The effect of this competition is such that each network is forced to offer a combination of fixed and usage fee that maximises customers' welfare subject to the constraint that the networks break even. This implies, in turn, that usage is charged at marginal cost and the fixed fee is set in such a way that each network just breaks even. Since each network does not earn any revenue from outgoing calls, the only other source of revenue is the revenue each network earns from terminating traffic from the fixed network. If the fixed cost per customer is $100 and the revenue from terminating traffic from the fixed network is $20 then the fixed fee per mobile customer is competed down to $80.

Suppose now that the number of mobile customers depends on the fixed fee. For example, it might be that there are 100 extra mobile customers for every $10 reduction in the fixed fee. In this case, every $10 increase in the revenue from terminating traffic per customer corresponds to a reduction in the fixed fee of $10 and therefore 100 extra mobile customers.

Now consider the perspective of the regulator. The regulator can increase the termination revenue of mobile networks by increasing the price for terminating a call on a mobile network. If there are 1 000 calls to each mobile customer, then an increase in mobile termination charge of just 1 cent increases termination revenue per customer by $10, increasing the number of mobile customers by 100.

Does it make sense to increase the termination revenue in this way? That depends on whether or not the mobile termination charge is passed on to customers of the fixed network in the form of higher fixed to mobile call charges. If the termination charge is passed on to customers, then there is a trade off – higher termination charges increase the number of mobile customers that fixed network customers can call, but raises the cost and reduces the demand for calling any one of them.

The sequence of events is then:

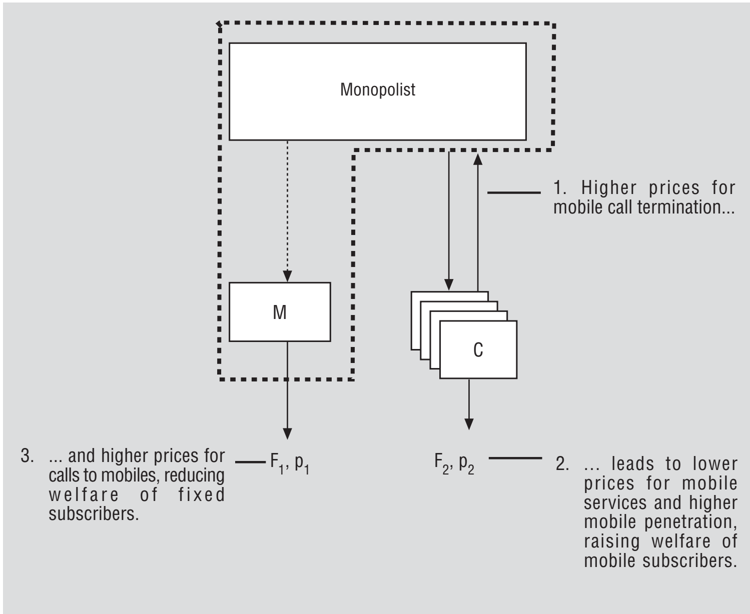
a) higher termination charge higher termination revenue lower mobile fixed charges more mobile subscribers; and

b) higher termination charge higher fixed to mobile prices fewer fixed to mobile calls per mobile subscriber

This issues is discussed in more detail in the appendix.

As usual, the way to resolve this conflict is to look for another "instrument". One such instrument might be simply to break the link between the fixed-to-mobile price and the termination charge by setting the fixed-to-mobile retail price independently of the mobile termination charge. If these prices were set

Figure 6. **Effects of higher mobile termination charges**



Source: OECD.

separately, the fixed-to-mobile charge could be set efficiently, relative to "true" marginal cost and the termination charge would then simply have the effect of shifting profits from the fixed to the mobile network, in order to balance mobile penetration against the harmful effects of price-cost mark-ups.

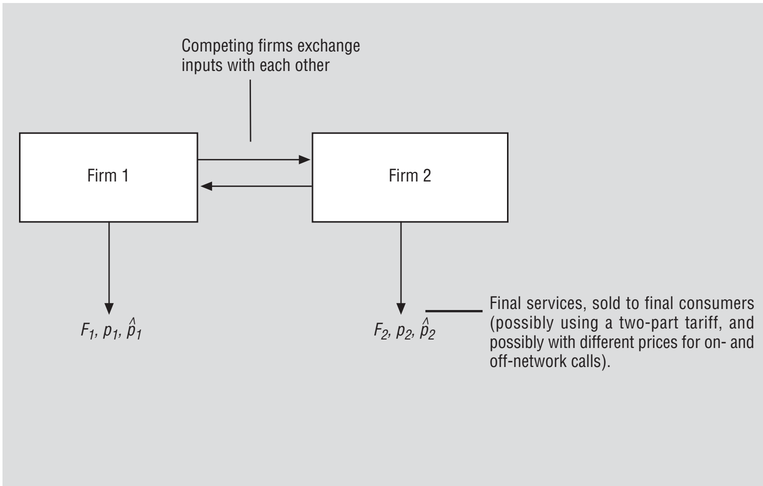
A more realistic alternative solution is to use a two-part access price, with, for example, a "fixed" component based on the number of mobile subscribers and a "variable" component based on the number of fixed-to-mobile calls. With such a structure, the efficient access price would then have the variable component (the mobile termination charge) equal to the marginal cost of mobile termination. The fixed would be chosen to balance mobile penetration against the harmful effects of price-cost mark-ups.

4. Two-way access pricing: interconnection of two competing networks

In the previous section we focused on a model in which there was a central monopoly firm exchanging inputs with a number of downstream competing firms. In that model the prices of the monopoly and the price at which the downstream firms sold their inputs to the monopolist were regulated and the prices of the downstream firms determined by competition.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Figure 7. **Market structure of a two-way access problem with competing firms exchanging inputs**



Source: OECD.

In this section, we consider a model in which two firms both compete with each other and exchange essential inputs. In this model, all retail prices are determined by competition. The regulator's role is limited to determining the access prices – the prices at which the firms exchange inputs.

Suppose, then, that we have two competing networks. These networks compete over a fixed set of consumers. All of the consumers choose one or other of the networks – there is therefore no role for enhancing penetration overall. As before, consumers care only about their own welfare and not that of their callers – nor do they receive a utility from calls to them. Consumers choose the network to which they will subscribe according to a Hotelling model of product differentiation. The two networks are assumed to offer services, which are differentiated in the eyes of consumers. As a result, one network can maintain higher prices than the other without losing all its customers.

As in the previous section, increasing the amount a network charges for terminating calls from the other network can have two effects – it may both increase the profits from termination (which can be used to lower other prices, enhancing demand for the network) and (to the extent that the increased termination charge is passed on to customers of the other network) it may increase the rival's prices – which itself, in turn, increases demand for the first network's services. For this reason, if it is allowed to do so, each network has a strong unilateral incentive to increase its own termination charge to the monopoly price. This is an illustration of the principle mentioned earlier – that if allowed to set a termination charge unilaterally and non-co-operatively, each network will choose a termination charge that is inefficiently high.

For this reason, we will focus on the case where the termination charge is restricted to be reciprocal (*i.e.*, the charge for termination is the same in each direction independent of the network). We are primarily interested in answering the following questions:

1. First, what is the effect of the termination charge on competition between the networks? Does raising the termination charge lead to a net flow of calls from one network to the other? Does a high termination charge favour the larger network? Does the larger network, other things equal, prefer a higher termination charge?
2. Second, will the two networks be able to agree on a common termination charge and if so, at what level?

In all of the models that we will look at below, raising the termination charge above marginal cost has the following effects:

1. First, as long as both networks have some positive size, raising the termination charge raises the cost of terminating inter-network calls and therefore raises the average cost of call termination. The average cost of call termination is raised by a larger amount for the smaller network (for which more calls terminate on the other network) than for the larger network.
2. The higher average cost of call termination may (depending on the assumptions of the model) directly affect the price of calls, which, in turn, affects the net imbalance in flows of calls across the networks. For example, in Box 4, an increase in the termination charge leads to an increase in the price of calls – and the increase is larger for the smaller network.
3. Increasing the termination charge also affects the revenue from call termination. The total effect on call termination revenue depends on the combined effects of the increased revenue per terminated call and the effect on the net imbalance in flows of calls across the networks.
4. Finally, increasing the termination charge may also have an indirect effect on other prices, such as the fixed component of two-part charges. These may go up or down depending on the combined effects of higher costs on the one hand and higher revenue from termination on the other.

The overall effect depends on the nature of competition between the two networks, such as whether they compete in simple linear prices or in two-part charges. For example, consider the following simple model – the two networks compete in two part tariffs, with a fixed charge and a per call charge. Suppose that a new entrant network must offer both a lower fixed charge and a lower per call charge in order to attract customers away from the incumbent's network. In this scenario a high termination charge disadvantages the entrant in two ways: First, it raises the average cost of termination for the entrant relative to the incumbent. Second, the entrant's lower per call charge will mean that there will be a net outflow of calls from the entrant's network to the incumbent's network. This net outflow, combined with an above-cost termination charge, leads to a net payment from the entrant to the incumbent. By assumption, the entrant cannot recuperate its higher costs in the form of a higher fixed fee.

This simple example suggests that dominant incumbent operators will always prefer a high termination charge, while new entrant operators will prefer a low or zero termination charge. However, this simple model is incomplete in certain ways. In particular, why should it be that the entrant has to offer both a lower fixed charge and a lower per call charge? Is it not the case that some consumers would be willing to pay a slightly higher per call charge if that also implied a lower fixed charge, or *vice versa*?

The development of more carefully specified models is left to the appendix. The results from the appendix can be summarised as follows. In a model of two competing firms which exchange inputs, without concerns over promoting penetration, the efficient access price is just equal to the marginal cost of termination in each direction. Even if we insist that the termination charges must be the same in each direction, there is no assurance that private negotiations between the networks will lead to the efficient access prices. In fact, if the networks are not able to use two-part pricing in their final prices, then they will have an incentive to collude to raise the access prices in order to raise final prices.

On the other hand, when the two firms compete in both fixed and variable charges, and demand is symmetric, there are no net flows of calls between the two networks in equilibrium, so the networks are indifferent as to the level of the common access charge. If demand is not perfectly symmetric and the access charge is not equal to marginal cost, there is a net transfer from the larger to the smaller network. As a result, the larger network always prefers the access charge to be equal to marginal cost. This suggests that it may be possible, in some circumstances, to allow the larger network to set the common access charge unilaterally.[29]

These models may also apply to the interconnection of Internet networks. In practice, major operators of Internet networks have agreed to interconnect with one another on a bill-and-keep basis. This may be due to the desire to enhance penetration (and, as we have seen, in a model where penetration is not an issue, network operators are indifferent to the interconnection charge when they compete using two-part charges). It is possible that, in the future, one of these networks will seek to change the interconnection charging structure, perhaps by introducing different prices for on- and off-network communications.

5. Conclusion

This chapter has sought to derive a set of insights from the theory of access pricing. We have seen how it is possible to distinguish one-way and two-way access pricing. The principles for setting one-way access prices follow directly from the principles of regulation of a natural monopoly. We emphasised, in particular, the importance, from the competition perspective, of ensuring that any price discrimination in access prices is reflected in final prices and *vice versa* . These principles clearly have direct application in the regulation of access in other public utility sectors, apart from telecommunications.

In the case of two-way access pricing, economic models tend to be much more specialised, more focused on the telecommunications industry and more sensitive to particular assumptions. Rather than drawing general principles, a range of models were presented showing the outcomes that might be expected and highlighting the sensitivity of the conclusions to assumptions about the particular form of competition between the networks.

The next chapter of this report seeks to further deepen our understanding of access pricing by bringing this theory alongside the practice of access regulation as it is carried out in OECD countries.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Appendix to Chapter 1

Efficient pricing of substitute goods

Suppose that two goods are substitutes or complements, and the price of the second good is held fixed and different from marginal cost. What price for the first good maximises welfare, given that the price for the second good is not equal to marginal cost?

Consider selling another unit of the first good and simultaneously adding or withdrawing units of the second good so as to maintain its price constant. If the two goods are substitutes then selling one more unit of good 1 would normally lower the price of good 2, so we have to sell fewer units of good 2 to keep the price of good 2 constant. Conversely, if the two goods are complements then selling one more unit of good 1 will imply selling more units of good 2, holding the price of good 2. Let σ be the reduction in the number of extra units of good 2 that must to be sold in order to keep the price of good 2 constant ($\sigma > 0$ for substitutes, $\sigma < 0$ for complements and $\sigma = 0$ when the two goods are neither substitutes nor complements). Let p_1, p_2 and c_1, c_2 be the prices and marginal costs of goods 1 and 2 respectively. Then, selling an extra unit of good 1 increases overall welfare by $p_1 - c_1$ and selling σ fewer units of good 2 decreases welfare by σ $(p2 - c2)$. The overall effect on welfare is then: $p_1 - c_1 \angle \sigma\ (p2 - c2)$. In order for p_1 to be the welfare-maximising price it must be that this expression is equal to zero, or, equivalently:

$$p_1 = c_1 + \sigma (p2 - c2)$$

In other words, if the price of good 2 is different from marginal cost then the price of good 1 should also be different from marginal cost and in the same direction for substitutes and opposite direction for complements. In the special case of perfect substitutes $\sigma = 1$ and identical marginal costs, this reduces to: $p_1 = p_2$.

Access pricing with substitute goods and the ECPR

The previous section showed that if p_1, p_2 and c_1, c_2 be the prices and marginal costs of goods 1 and 2 respectively and if σ is the reduction in the number of extra units of good 2 that need to be sold in order to keep the price of good 2 constant, then the correct relationship between the prices is given by:

$p_1 = c_1 + \sigma(p_2 - c_2)$

In the context of access pricing, equation (3) in the section below shows that if is c_0 the marginal cost of the essential input, and if producing one unit of goods 1 and 2 requires one unit of the essential input, then the efficient access price is:

$a_1 = c_0 + \sigma(p_2 - c_0 - c_2)$

Which can be re-written as follows:

$a_1 = (1 - \sigma)c_0 + \sigma(p_2 - c_2)$

This way of writing the access price makes clear the relationship between different approaches to access pricing. When goods 1 and 2 are neither substitutes nor complements ($\sigma = 0$), the access price is just given by marginal cost. When goods are perfect substitutes ($\sigma = 1$) the access price is given by the simple ECPR formula (the final price less the monopolists marginal cost in the competitive activity).

$a_1 = p_2 - c_2$

We can extend this result slightly to derive another useful formula. Suppose that the access input is used to produce a service that is a perfect substitute for the service of the monopolist and suppose that the monopolist discriminates in the prices of its final service. Let $\hat{p}$ and $\hat{c}$ be the price and marginal cost of the monopolist's final service. When the monopolist grants access it may nevertheless recover some revenue directly from the final customer and it may still incur some costs. Let *p* and be the price and marginal cost incurred by the monopolist in the provision of the service once access has been granted. Let *a* be the access price. In order to preserve the contribution to fixed costs, the contribution before and after granting access must be the same. This implies that:

$a = (p - c) - (\hat{p} - \hat{c})$ *or, equivalently,* $a = (p - \hat{p}) - (c - \hat{c})$

This can be viewed as an extension of the simple ECPR. The key point to note here is that this expression must be applied to all of the differentiated parts of the monopolist's final price. In other words, if the monopolist discriminates by charging different prices for two different classes of customers, then this expression must be applied to each of those different final prices. If the monopolist uses a two-part tariff in its final prices, then the above expression must be applied to both the fixed and the usage price.

Ramsey pricing and access prices

Suppose that we have a monopolist that produces two services labelled 1, 2. When the prices of the two goods are p_1 and p_2, the demand curves for these products are q_1 (p_1, p_2) and q_2 (p_1, p_2), respectively. The consumers' surplus is defined to be $V(p_1, p_2)$ where $\frac{\partial V}{\partial p_i} \equiv -q_i(p_1, p_2)$.

The marginal cost of producing the goods is c_1 and c_2, so the profit in producing good *i* is $\pi_i(p_1, p_2) \equiv (p_i - c_i)q_i(p_1, p_2)$. The social welfare function $W(P_1, P_2)$ is the sum of consumers' surplus and producers' surplus (*i.e.*, the profit function).

What prices maximise social welfare, in the absence of any other constraints? Let p^* be the social welfare maximising prices. By the first order conditions:

$$\frac{\partial W}{\partial p_i}(p_1^*, p_2^*) = 0 \Leftrightarrow (p_1^* - c_1)\frac{\partial q_1}{\partial p_i} + (p_2^* - c_2)\frac{\partial q_2}{\partial p_i} = 0$$

Which is satisfied when

$$p_1^* = c_1 \text{ and } p_2^* = c_2 \qquad \ldots(1)$$

In other words, social welfare is maximised by choosing all prices equal to marginal cost.[30]

Suppose now that the price of the second good is held fixed and different from its marginal cost, *i.e.*, $p_2 = \overline{p} \neq c_2$. What price for the first good maximises social welfare. From the first order condition, this price must satisfy:

$$(p_1^* - c_1)\frac{\partial q_1}{\partial p_1} + (\overline{p} - c_2)\frac{\partial q_2}{\partial p_1} = 0$$

So, in other words:

$$p_1^* = c_1 + \sigma(\overline{p} - c_2) \text{ where } \sigma = -\frac{\partial q_2}{\partial p_1} / \frac{\partial q_1}{\partial p_1} \qquad \ldots(2)$$

σ is a ratio which reflects how much a change in the price of the first good changes demand for the second good relative to changes in demand for the first good. If two goods are independent then $\sigma = 0$, so the second good should be priced at marginal cost. If two goods are perfect substitutes, $\sigma = 1$.

Now suppose that the first good is an input which is sold both to a downstream competitive industry and to a downstream affiliate of the monopolist itself. The marginal cost of producing the access good will be c_0. The marginal cost of the converting this access good into a final good will be c_C for the downstream competitive industry and c_M for the monopolist. Let the price of the access good be a_1. The price of the final goods will be p_c and p_M for the output of the downstream competitive and monopolist industries respectively.

Since the downstream industry is competitive, we have that price is equal to marginal cost, *i.e.*,

$$p_C = c_C + a_1$$

Now, using the results above we know that in the event that the final product of the monopolist is fixed away from cost (*i.e.*, $p_M = \overline{p} \neq c_M + c_0$) then

$$a_1 = p_C - c_C = c_0 + c_C + \sigma(\overline{p} - c_0 - c_M) - c_C$$

So that

$$a_1 = c_0 + \sigma(\overline{p} - c_0 - c_M) \qquad \text{...(3)}$$

where $\sigma = -\frac{\partial q_M}{\partial p_C} / \frac{\partial q_C}{\partial p_C}$. In other words, the efficient access price is equal to the marginal cost of access plus a term reflecting the degree of substitution between the final price of the monopolist and the competitive industry. This is one (generalised) way of stating the Efficient Component Pricing Rule.

In the special case of perfect substitutes, ($\sigma = 1$), we have the special result:

$$a_1 = \overline{p} - c_M \qquad \text{...(4)}$$

I.e., the access price is just the monopolists final price less the monopolists marginal cost of converting a unit of the input into a final good. We might refer to as the simple ECPR.

Suppose now that the total quantity available of the input is fixed. In other words:

$q_M(p_C, p_M) + q_C(p_C, p_M) = \overline{q}$, where $\overline{q}$ is a constant. In this case, we have that,

$\frac{\partial q_M}{\partial p_C} + \frac{\partial q_C}{\partial p_C} = 0$ so $\sigma = 1$. In other words, the expression $a_1 = \overline{p} - c_M$ gives the correct access price when the total quantity of the available input is fixed, even when the final service of the competitors is not a substitute for the final service of the monopoly.

Now let's return to our original notation. Suppose that pricing at marginal cost is not feasible due to the presence of fixed costs which cannot be subsidised from outside the industry. What is the efficient set of prices given the requirement that the firm must at least break even? The problem, then, is to maximise social welfare subject to the constraint that the monopolist makes non-negative profits. The lagrangian for this problem can be written:

$$L = W(p_1, p_2) - \lambda\Pi(p_1, p_2)$$

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

The prices which maximise this expression must satisfy:

$\frac{\partial L}{\partial p_i} = 0 \Leftrightarrow (1-\lambda)((p_1 - c_1)q_{1i} + (p_2 - c_2)q_{2i}) = \lambda q_1$ where $q_{ji} = \frac{\partial q_j}{\partial p_i}$, which can be written as: $\frac{p_i - c_i}{p_i} = \frac{k}{\eta_i}$ where $k = \frac{\lambda}{1-\lambda}$ and $\eta_1 = \frac{p_1(q_{11}q_{22} - q_{12}q_{21})}{(q_1 q_{22} - q_2 q_{21})}$ and

and similarly for η_2 …(5)

Mobile termination

Suppose there are *N* mobile subscribers and each receives Q fixed-to-mobile calls. Let *P(a)* be the price of a fixed-to-mobile call when the termination charge is *a* and let $P'(a)$ be the first derivative of *P(a)* with respect to a.

Consider the effect of raising the termination revenue by $1. This increase in termination revenue is passed on to mobile subscribers in the form of lower mobile fixed charges, raising the welfare of each mobile subscriber by exactly $1, contributing $*N* to overall welfare. It also raises the number of mobile subscribers by a certain amount, increasing the welfare of fixed subscribers, by an amount *A*, say. However, raising termination revenue by $1 implies increasing the price of fixed-to-mobile calls, reducing the welfare of fixed subscribers. If the price is raised by ΔP, the welfare of fixed subscribers is reduced by exactly *NQ ΔP*. If the termination charge was previously set at marginal cost, raising the termination revenue by $1 requires raising the access charge by $P'(a)/Q$. In other words, a $1 increase in the termination revenue reduces the welfare of fixed subscribers by exactly $P'(a)N$. Adding together all of these effects, we find that an increase in $1 in termination revenue (when the access charge is set at marginal cost) increases total welfare by $A - P'(a)N + N$.

We can use this expression to demonstrate the results in the text. Suppose that increasing the termination revenue increases mobile penetration (*i.e.*, $A > 0$) and that the fixed-to-mobile price is equal to marginal cost (*i.e.* $P'(a) = 1$). Then the expression above says that raising the access charge above marginal cost increases overall welfare by the amount *A*. The optimal access charge is therefore greater than marginal cost.

On the other hand, if termination revenue has no effect on mobile termination ($A = 0$) and the fixed-to-mobile price is marked-up above marginal cost (*i.e.*, $P'(a) > 1$) then the above expression says that raising the access charge above marginal cost lowers overall welfare. The optimal access charge in this case is below marginal cost $N(1 - P'(a)) < 0$.

A closer look at competing networks with two-part tariffs

Suppose that we have two networks, labelled A and B. which compete using a two part tariff. Each network offers its customers a pair of prices (*e.g.*, f_A, p_A) with a fixed component and a per call component. The marginal cost of originating and terminating calls is just *c*.

The prices offered by each network determine each network's market share. Let n_A and n_B be the number of subscribers to each network, where $n_A + n_B = 1$.

If we hold constant the utility of subscribers, we find that the per call price which maximises profits is simply the price equal to perceived marginal cost, *i.e.*, $p_A = n_A(c+c) + n_B(a+c) = 2c + (a-c)n_B$.

Suppose that the two networks have very different sizes. For example, suppose that network A is very much larger than network B, so that $n_A = 0.9$ and $n_B = 0.1$ and suppose the marginal cost of call origination and termination is $1 ($c = 1$). Then $p_A = 0.9 \times 2 + 0.1(a+1) = 1.9 + 0.1a$ while $p_B = 0.1 \times 2 + 0.9(a+1) = 1.1 + 0.9a$. It is clear that the price of the larger network is much less sensitive to the access charge than the price of the smaller network. For a high access charge, say $a = 2$, $p_A = 2.1$ and $p_B = 2.9$.

With an access charge above cost, there is a net outflow of above-cost calls from the larger to the smaller network and therefore a net transfer payment in favour of the smaller network. When the access charge is below cost, say $a = 0$, $p_A = 1.9$ and $p_B = 1.1$, so there is a net outflow of below-cost calls from the smaller network to the larger network and so, again, there is, in effect, a transfer payment in favour of the smaller network.

In the case of symmetric demand for the networks, in any equilibrium in which $p_A = p_B$, there is no net flow of calls from one network to the other. One consequence is that the profits of each firm in equilibrium do not depend on the access charge – these firms are indifferent as to the regulator's choice of the access charge.

Models of two-way networks

Competition in simple linear prices

Suppose, first, that the two networks do not impose a fixed charge and are not allowed to price discriminate between on and off-network calls. In other words, suppose that the two networks compete in simple linear (per call) prices.

In this case both networks may prefer high termination charges, even when one network is smaller than the other one. To see this, let's suppose that at an equilibrium the prices of each network are the same. This implies, in turn, that there is no net flow of calls between the networks regardless of the relative size of the two networks. An increase in the termination charge therefore has no effect on net termination revenue (which is zero). On the other hand, an increase in the termination charge increases the marginal cost of termination and therefore leads to an increase in the price for calls that each network chooses. As a result, each network prefers setting the access charge high enough to induce the monopoly price for calls.

If one of the networks tries to deviate from this equilibrium by cutting its price, it will simultaneously increase its market share and induce a net outflow of calls to the other network. If the access charge is high enough the benefits of increased market share are outweighed by the increased termination payments, so neither network has an incentive to deviate slightly. (On the other hand, if the networks are close substitutes a more substantial deviation may allow one network to capture all of the market and thereby all of the profits – rather than sharing the market with its rival – so this collusive outcome may not be possible when the two networks are close substitutes.)

To summarise, when networks compete in simple linear prices a high termination charge is preferred by both networks as it allows both networks to charge the monopoly price. Carter and Wright (2001) summarise this:

> "Symmetric local networks will set the common access charge above the cost of providing access to lessen retail price competition. By agreeing to high interconnection charges, firms reduce the incentive to undercut each other in an endeavour to increase market share. If either firm lowers its retail price it will face a net outflow of calls which, given sufficiently high interconnection charges, will reduce its profits. In this way, competition in the retail market can be undermined by collusion over the access charge."[31]

Carter and Wright (1999) show that this result also applies in the case of unequal-sized firms, in a model of brand-loyalty.

Competition in two-part charges without differentiation between on and off-network calls

Now let's consider the more realistic case in which the networks can charge a fixed fee in addition to a usage fee. We will also assume for the moment that price discrimination between on- and off-network calls is prohibited and (as before) the termination charge must be reciprocal.

In this case we find that each firm will choose to set the per call price equal to the marginal cost of calls. But the marginal cost of calls depends, in turn, on the termination charge. Since, for the smaller network, more of its calls are cross-network, a high termination charge will increase the per call prices of the smaller network more than the larger network. This is shown in the appendix.

A high termination charge does not necessarily act against the interests of the smaller network, however. Since the smaller network has higher usage prices, there will be a net flow of calls towards the smaller network. The smaller network earns revenue from terminating these calls. In fact, the smaller network may actively prefer a high interconnection charge precisely because of the termination revenue that it receives.

Carter and Wright (2001) have explored this issue. They find that the larger network would always prefer an access charge equal to marginal cost. If the access charge differs from marginal cost (either higher or lower), the larger network loses revenue. To see this, observe that when the access charge is above marginal cost, subscribers to the smaller network make fewer calls than subscribers of the larger network, so there is a net inflow of calls to the smaller network and therefore a net payment of termination charges in the direction of the smaller network. When the access charge is below marginal cost, subscribers to the smaller network make more calls than subscribers of the larger network, so there is a net inflow of calls to the larger network. But since the termination charge is below marginal cost the larger network makes a loss terminating those calls.

For the same reason, the smaller network will, under some circumstances, prefer an access charge different from marginal cost. Carter and Wright propose that one solution to the access pricing problem in this context might be simply to allow the larger network to unilaterally choose the common termination price:

> "Our model suggests a very simply principle can achieve desirable outcomes for local exchange carrier interconnection – let the incumbent pick the access price and apply it reciprocally. The incumbent will not want to pick an access price above cost, since this will lead to its smaller rivals competing with high usage prices but lower rentals. The end result will be an outflow of calls from the larger firm and an access deficit. Similarly, the incumbent will not want to pick an access price below cost, since this will lead to its smaller rivals competing with low usage prices and higher rentals. The end result will be an outflow of calls to the larger firm, which given the below cost access prices will also imply an access deficit."[32]

In the case when demand for the networks is symmetric, it turns out that each network is indifferent in equilibrium to the level of the access charge. In a symmetric equilibrium, the usage prices are the same on each network. As a result, there is no net flow of calls between the networks, so the profits earned in equilibrium do not depend on the access charge. The networks are completely indifferent to the level of the access charge.[33]

Box 5. **A closer look at competing networks with two part tariffs**

Suppose that we have two networks with very different sizes. Network A is very much larger than network B. We will assume that the price of calls is always equal to marginal cost (it is possible to show that when two networks compete in two part tariffs they have an incentive to set the price of calls equal to marginal cost in equilibrium). The two networks have the same marginal cost of origination and termination and charge a reciprocal termination charge for terminating calls from the other network. Each subscriber calls every other subscriber with equal probability.

If the termination charge is equal to marginal cost then (since the marginal costs are assumed to be the same on both networks) it is clear that the prices of the two networks will be the same. If the termination charge is above the marginal cost of termination what effect does this have on the prices charged by each network?

Suppose that network A has 90% of the customers and network B only 10%. For any one customer of network A, there is a probability of 90% that the call will terminate in network A and a probability 10% that the call will terminate in network B. For calls that terminate in network B, network A must pay the above cost termination charge. In contrast, for network B there is a 90% chance that a call will terminate in network A and incur the above cost termination charge. As a result, the prices of network B are much more sensitive to the termination charge than are prices of network A. A small increase in the termination charge raises both prices of network A and B, but it raises the prices of network B by much more.

Since the prices of network B are raised by more than the prices of network A, an effect of raising the termination charge above marginal cost is that it reduces demand for calls on network B more than network A, leading to a net flow of calls from network A to network B. This net imbalance of calls is a source of revenue for the smaller network, allowing it to reduce its other charges.

If the termination charge is reduced below marginal cost, the price of network B reduces faster than the price of network A. Demand for calls increases on network B faster than on network A. As a result there is a net flow of calls to the larger network. Since the termination charge is below marginal cost the larger network makes a loss on the termination of these calls, forcing it to raise its other charges.

Competition in two-part charges with differentiation between on- and off-network calls

What if we allow firms to price discriminate between on-and off-network calls? In this case it is possible to show that in a symmetric equilibrium (if one exists) the two networks would jointly prefer a termination charge which is *below cost*. The reason is that a below-cost termination charge softens competition between the networks for customers. If termination charges are below cost then an attempt by one network to expand raises its costs – as now a higher proportion of calls remain on its network and must be terminated at cost. In effect, when the termination charge is below cost each network wants to lose customers rather than gain them. This has the effect of softening competition for customers, allowing the networks to raise other charges.

Armstrong (2001) summarises this as follows. When the termination is charged below cost:

> "… it is cheaper for subscribers to call people on their own network, and so, all else equal subscribers prefer to belong to the *smaller* network. This means that the market exhibits 'negative network externalities', and firms have little incentive to compete aggressively for subscribers. In effect, by means of a suitable choice of termination charge firms can decide whether to have a market with positive network externalities [termination charge above marginal cost], no network externalities [termination charge equal to marginal cost] or negative network externalities [termination charge below marginal cost]. Because of its softening effect on competition, it is mutually profitable for firms to choose the third option."[34]

This last result is likely to depend strongly on the assumption that demand is symmetric. If demand is asymmetric, it seems likely that the larger network would prefer a higher access charge, as that increases the price for inter-network calls. Since smaller networks have more inter-network calls, that would, in turn, increase the demand for the larger network.

All of the above models share certain common, rather artificial, assumptions. Relaxing these assumptions may change the predictions of the models. For example:

1. In the above models, individual subscribers were assumed to be identical and to have identical demands for calling all other subscribers. In practice certain subscribers receive more calls than they generate ("call sinks") or *vice versa*. If the access charge is above the marginal cost of termination, networks have an incentive to generate more net termination traffic in their favour. They can do this by signing up "call sinks" – *i.e.*, individual subscribers who receive many more calls than they make. The "modem pools" of Internet service providers are a classic example. A company can also increase its net termination traffic by refusing to sign up subscribers with a large net outflow

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

of calls such as telemarketers. As long as access prices are above marginal cost and as long as it is possible to discriminate between final customers, telecommunications companies will compete more fiercely for final customers which have a net inflow of calls and will tend to shun customers with a large net outflow of calls.

2. In the above models, all subscribers always chose one network or the other – in other words, penetration was 100% so increasing penetration was not a public policy concern. A possible extension of the above models would allow for the effects on penetration. We might expect that if two networks differ in the sensitivity of their subscribers to their charges, then it might be efficient for the network with the relatively insensitive customers to subsidise the other network, to enhance penetration. Ideally, this would be done through a fixed charge (rather than a usage charge), such as a charge per subscriber.
3. All the above models featured only two competing networks – what happens if there are multiple competing networks? Is it still the case that we can rely on the largest network to set the common interconnection charge unilaterally, as proposed earlier?
4. Another possible extension of the models would be to look at the effects of termination on barriers to entry. It seems intuitively likely that high termination charges would facilitate entry in the case of linear final prices and no price discrimination between networks but would hinder entry in the case of two-part charges and price-discrimination across networks. Intuitively, the reason is that with no discrimination between on and off-network calls, there is no disadvantage to a network from being small, whereas if it costs more to call another network there is a sizeable disadvantage to being a small network, which may increase the costs of entry.

Notes

1. In order for competition to operate, however, it is essential that the total or combined price to the end-user depends directly on the payment to the competitive component (in this case the ISP) and that the competitive firms have control over the size of the payment to the competitive component.
2. Usually defined as the sum of consumers' surplus and producers' surplus (less the social cost of raising any necessary subsidies).
3. This issue is discussed further in the last chapter of this report.
4. In essence, I am assuming here that other public policy objectives must be satisfied before the allocative efficiency objective. Strictly speaking, since pursuing these other objectives involves a loss of allocative efficiency (distorting prices away from marginal cost) these other objectives should be traded off against the allocative efficiency objective, which would typically involve some loss of both the other objective (such as geographic uniformity) and allocative efficiency.
5. One common rationale for competition, as well as for certain types of access pricing rules, is that limiting the discretion of regulators to redistribute industry benefits may lead to greater efficiency.
6. Strictly speaking, the case of perfect substitutes is a special case in which it is difficult to speak of two separate goods. If two goods are perfect substitutes only the lower of the two prices matters. If the two goods have a different marginal cost then only the good with the lower marginal cost should be produced and consumed.
7. This principle also shows that, other things equal, not being able to tax related goods outside the industry constrains the efficiency of revenue collection.
8. Specifically, if marginal cost pricing results in a deficit, raising the funds to cover that deficit from outside the sector (perhaps through general taxation) will typically involve some other inefficiency. Taking into account those other losses will generally imply some degree of contribution towards the deficit from revenue raised within the sector. In addition, full cost recovery allows some test of whether or not the service is socially desirable at all. A service which covers its marginal cost may not generate enough total surplus to justify its existence, whereas a service which covers its total cost must generate enough benefits to consumers to justify its existence.
9. If entrants can produce the same services at the same cost as the monopolist the service is competitive and is therefore not an essential input.
10. This is just a necessary condition – entry might still be deterred, even with revenue above stand-alone cost. In the extreme case where there are no barriers to entry for *any* service or group of services (*i.e.*, new entry occurs instantly whenever the revenue from a service or group of services exceeds the stand-alone cost of providing that service – also known as perfect contestability) then it can be shown that revenue from *any* service (or group of services) should never be lower than the incremental cost of that service.

11. See Lewis and Sappington (1988a, 1988b).
12. This conclusion rests on the assumption that the regulated prices do not affect the profitability of any unregulated services produced by the regulated firm. Otherwise the firm may distort prices under the cap to increase the profits derived from uncapped services.
13. One of the submissions to this chapter asks how, if the downstream industry is competitive, the final price can be above marginal cost. The answer is that the assumption that the downstream industry is competitive ensures only that the final price is equal to the marginal cost of the downstream services. The access price is one component of the marginal cost of the downstream services. If the access price is above the marginal cost of the access service the final price will be above the marginal cost of the combined upstream and downstream services.
14. This is demonstrated in the appendix.
15. Possibly including a contribution to fixed costs as set out in principle 3.
16. See OECD (2002), page 133.
17. The proportionality restriction may limit efficiency in comparison with the case in which the level of fixed charges paid by a competitor may depend on additional customer characteristics.
18. Of course, the price discrimination of the incumbent is not always necessarily efficient. The failure to price discriminate at the access level may actually assist to eliminate inefficient forms of price discrimination.
19. Commission Recommendation on Interconnection in a liberalised telecommunications market, C(97) 3148.
20. The assumption of fixed input proportions is not a trivial one with respect to this statement. If, instead, the monopolist under the global price cap is only able to base its charges to downstream competitors on their demands for the access component, but not their sales of final services, competitors will inefficiently substitute away from the monopoly input if they can.
21. This assumes that the global price cap is truly global, *i.e.*, products in related markets (substitutes and complements) are also under the global price cap.
22. Note that this would only arise if the local loops sold to competitors are entirely in addition to (and do not compete in any way with) the local loops already provided by the incumbent. This might be the case if, for example, the entrant was granted bitstream access to a local loop while the incumbent continued to provide telephone service and assuming that the incumbent did not provide its own high-speed Internet service.
23. If the total number of local loops is fixed, then every local loop which is sold to a downstream firm corresponds to one less local loop which the incumbent uses to provide its own services. It is as though all possible alternative uses of the local loop are perfect substitutes, no matter what use those local loops are ultimately put. In addition, if the number of local loops is fixed, then the optimal solution is to ration them to whoever values them most. If the (single) access price that does this raises *too much* money and the excess can't go into the government's general revenues or be redistributed in lump sums, then some non-market rationing device would have to be found that attempted to do this, but the situation is different than the usual one of needing more revenue than can be efficiently gathered. Alternatively, if the single access price that rations the fixed supply raises too little money, access price discrimination may still be a useful way to proceed–

imagine doling out the lines to customers in order of the valuation placed on them and charging them epsilon (or greater) less than the value of their surplus.

24. This is also the presentation in Armstrong (2002).
25. Jerry Hausman, in a statement to the ACCC, states that "Empirical evidence from the US demonstrates that mobile subscribers place a value on incoming calls. In the US at this time mobile subscribers pay for incoming calls. Thus, anyone who receives incoming calls places a value on them since she is being charged for the call. If no value were received, the mobile subscriber would not give out their mobile number". Hausman (2000).
26. There are at least two situations where this might arise. The first is within a corporation which purchases mobile phones for its employees. The second is within a family where the cost of the fixed and mobile services are paid out of the common budget. In either case the price of calls to mobile services does affect the overall welfare of mobile customers and mobile customers do receive benefit from calls from fixed network subscribers.
27. In the case where the mobile termination charge is not passed on, *i.e.*, the fixed-to-mobile price is completely independent of the mobile termination charge – then the result we are trying to show here is clear: an increase in the mobile termination charge increases the profit of the downstream company with no effect on other prices or behaviour – so there is no theoretical limit on how high this price can be raised.
28. And, as mentioned earlier, in the theoretical case where the fixed-to-mobile price is independent of the mobile termination charge, there is no theoretical limit on the price the mobile network could charge.
29. If access prices are above marginal cost, competition will be intensified (and prices lower) for subscribers which produce a net inflow of calls (and *vice versa*). We might expect that with non-identical networks and some concern for promoting penetration, there would be a need for a two-part access charge.
30. Strictly speaking this would require verifying the second order conditions.
31. Carter and Wright (2001), page 2.
32. Carter and Wright (2001), page 12.
33. In addition, as before, if the two networks are close substitutes, a competitive outcome in which both firms co-exist in the market may not exist. As before, the reason is that if the two networks are close substitutes, one network may be able to capture the entire market through a relatively small cut in its prices. If the size of the cut in prices is small enough, the profit it earns from serving the entire market will exceed its profits when it shares the market with another firm.
34. Armstrong (2001), page 61. As before, this symmetric equilibrium may not always exist. But, in contrast with the earlier models, in this case the symmetric equilibrium fails to exist if the networks are sufficiently dissimilar. We have looked at the cases of linear pricing without price discrimination between on and off-network calls and two-part pricing with and without price discrimination. What about the case of linear pricing with price discrimination between on and off-network calls? In this case Laffont, Rey and Tirole (1998) show that when price discrimination is allowed and the networks are not close substitutes, the networks will choose to set the termination charge above cost. If networks are differentiated from each other, allowing differentiated prices for on- and off-network calls can improve welfare.

ISBN 92-64-10592-1
Acces pricing in Telecommunication

Chapter 2

The Practice of Access Pricing in Telecommunications

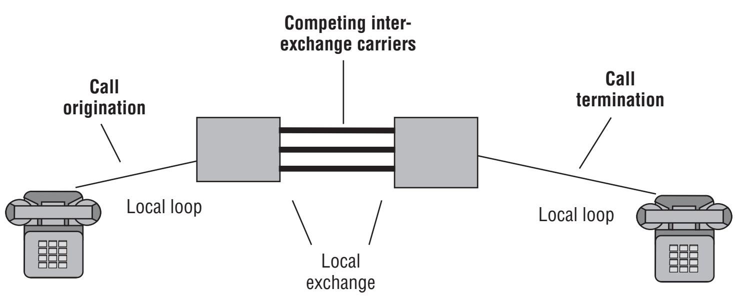
1. Introduction

The previous chapter set out the theory of efficient access pricing in some detail. In this chapter we seek to bring that theory along side the practice of access pricing in the telecommunications sector in OECD countries. We will focus on identifying which services should be regulated, on how the regulation of specific services interacts with the regulation of other services and on how the access prices relate to the prices of the corresponding retail or end-user services. Throughout we will seek to highlight where the theory and practice diverge.

The first step in the application of the previous chapter's theory to the telecommunications industry is the identification and classification of the access or "essential facility" problems that arise. As we saw in the previous chapter, we can distinguish one-way and two-way access problems. One way access problems arise when the provider of the competitive service needs access to a non-competitive service provided by another firm but not the other way around. In this report we will focus on the following one-way access problems in telecommunications:

1. access to call origination and termination at both ends of a call (for the purposes of providing competing end-to-end local, long-distance or international calls);[1]

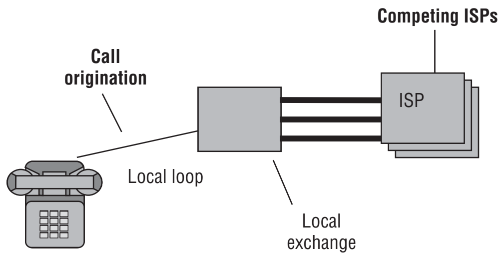
Figure 8. **One-way call origination and termination for voice calls**



Source: OECD.

2. access to call origination for the purposes of providing Internet access services;

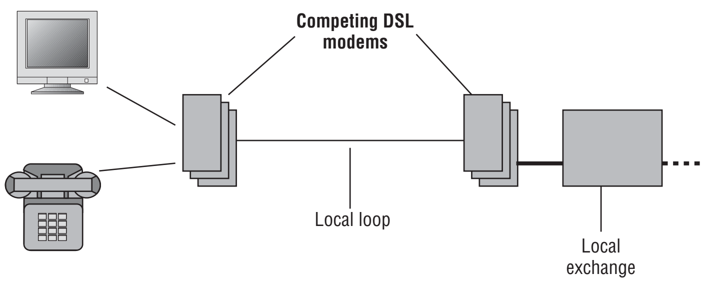
Figure 9. **One-way call origination for calls to Internet service providers**



Source: OECD.

3. access to the physical (usually copper wire) local loop facilities (*i.e.*, "local loop unbundling").

Figure 10. **One-way access to unbundled local loop**



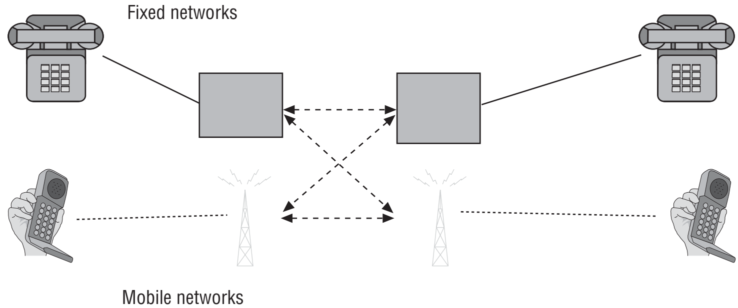
Source: OECD.

Two-way access problems arise when the owners of bottleneck facilities must purchase essential inputs from each other. This arises in telecommunications whenever subscribers attached to one network desire to communicate with subscribers attached to another network, as arises in the interconnection of two fixed, fixed and mobile or two mobile networks. This is illustrated in Figure 11.

There are a number of other telecommunications services which may qualify as "essential facilities" to which access is sometimes mandated but which will not be discussed further in this report, including:

1. transportation from the point of interconnection to the local exchange (in the case where the other network operator interconnects at a different level of the network);

Figure 11. **Two-way interconnection: fixed-fixed, fixed-mobile, mobile-fixed and mobile-mobile**



Source: OECD.

2. access to certain leased lines (specifically those lines which cannot be economically duplicated by competing operators and which are essential for competition);
3. access to rights of way (pipes and ducts), access to high sites (for antennas) or access to space within an incumbent's facilities (co-location);
4. access to numbering and electronic addressing resources;
5. access to roaming services on mobile networks;
6. access to conditional access services (*i.e.*, video encryption/decryption devices for interactive and pay television services).

2. Regulatory arbitrage and links between different access prices

Although we have carefully distinguished above between one-way and two-way access problems, in practice it is not always possible or desirable to distinguish between traffic that falls into the "one-way" or "two-way" category. In particular, as we will see, distinguishing between these two types of traffic may give rise to regulatory arbitrage (that is, a situation where an operator has an incentive to re-route or re-label traffic in order to benefit from lower tariffs) or may lead to a distortion of competition. Furthermore, it may not be possible to set different charges for call origination and termination, as this may distort competition for end-users that generate more calls than they receive (or *vice versa*).

Links between one-way and two-way call termination

Let's focus first on a potential link between one-way call termination and two-way call termination. The basic idea here is that whenever there are different charges for terminating different streams of traffic, there will be an incentive for telecommunications carriers to re-route or re-label their traffic to

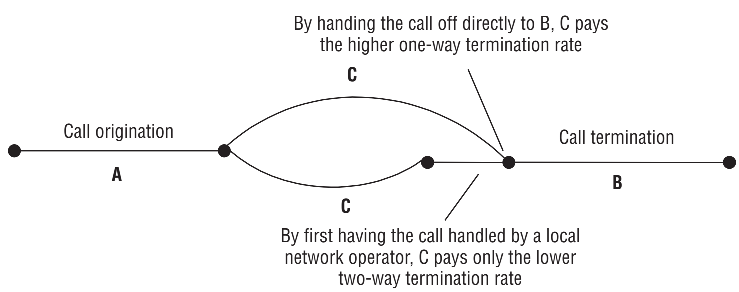
 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

benefit from the lower charges. Stricter regulatory intervention may be required to keep different streams of traffic apart.

To see this, suppose that two-way call termination is priced less than one-way call termination. For example, it might be that two-way call termination is not metered (as would occur under a "bill and keep" agreement) whereas the cost of terminating a long-distance call may be, say 4 cents per minute. Assume that the terminating company cannot determine the ultimate origin of a call, but can determine whether the company it is dealing with is licensed to provide local or long-distance calls.

In this circumstance there is a strong incentive for a long-distance operator to re-route its traffic through local network operators, so as to be able to interconnect with the incumbent local company and terminate calls at the lower rate for two-way termination. The difference in termination charges induces regulatory arbitrage – a long-distance company is induced to create a local network company that would possibly not otherwise exist to "arbitrage" the difference in the one-way and two-way termination charges.[2] This is illustrated in Figure 12:

Figure 12. **If two-way call termination is cheaper a long-distance operator may have an incentive to route its calls through a local operator**



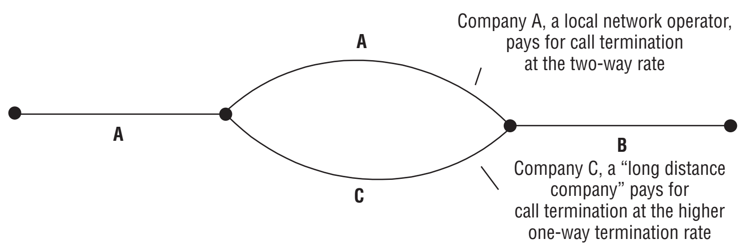
Source: OECD.

For example, in the US there have been complaints from IXCs about unfair competition from Internet telephony. On 19 April 2001 the FCC ruled that ISP Internet telephony traffic is interstate access traffic and not subject to the reciprocal compensation arrangements enjoyed by CLECs. However ISPs have been designated as "information access" and therefore do not pay the same termination and origination fees as IXC voice traffic. ISPs may therefore have a competitive advantage in the market for long-distance telephony.

A variant of this problem has also arisen in the mobile market. The charges for terminating international telephone calls were historically set using the international accounting rate system. For various reasons, that system did not distinguish between termination on fixed and mobile networks. As a result, in many cases the charge for terminating international traffic on a mobile network (which is here the "one-way" termination charge) is lower than the domestic fixed-to-mobile termination charge (which is here the "two-way" termination charge). This provides a strong incentive for domestic fixed carriers to seek to re-route their two-way termination traffic to make it appear like "one-way" traffic – that is, to re-route traffic to domestic mobile networks internationally. This is known as "tromboning". As an example, France Telecom pays 33 US cents to terminate traffic on a mobile network for a fixed-to-mobile call, but for terminating calls from neighbouring countries, France Telecom receives 8-9 cents per minute, of which it passes 5 cents to the mobile operator for terminating the call. There is a strong incentive for fixed network operators in France to route calls to French mobile networks via foreign fixed networks.[3]

In addition, a related regulatory distortion can arise if the regulatory regime differentiates traffic on the basis of the regulatory status of the company requesting termination. Companies that are allowed access to the lower charges have an advantage. In Figure 13, company A has an advantage because, having the status of a local network operator it can terminate calls at the lower two-way rate with B than can company C, which has the regulator status of a long-distance operator. To prevent this type of distortion, the regulator may be forced to impose line-of-business restraints on operators with a different regulatory status – the regulatory may be forced to prevent operators which are able to interconnect on the same basis as local operators (such as ISPs) from providing long-distance services.

Figure 13. **An integrated local and long-distance operator may have an advantage over a long-distance operator**



Source: OECD.

Links between one-way and two-way call origination

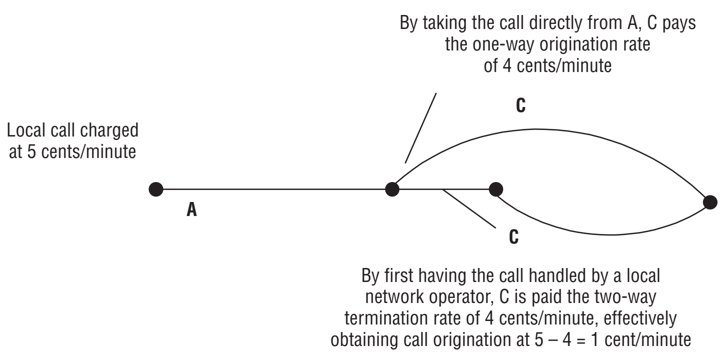
The same effects can also be seen on the side of call origination, although here the strategy that companies must employ to arbitrage differences in call origination charges appears at first a little more artificial. As before, if the

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

two-way tariff for call origination is lower than that for one-way call origination, there is an incentive for carriers which generate a lot of originating traffic to make their traffic appear as two-way traffic in order to receive advantageous regulatory treatment. The additional difficulty here is that the "two-way tariff for call origination" is not directly defined by the regulator, but is the difference between the retail price for a local call and the regulated tariff for two-way call termination.

To see this, suppose that one-way call origination and two-way call termination is charged at, say, 4 cents per minute, whereas a local call is charged to end-users at 5 cents per minute. In this case, if a long-distance operator connects directly to the incumbent it pays 4 cents per minute for one-way call origination. Alternatively, if it interconnects through a subsidiary local network, the incumbent collects an additional 5 cents for the local call, but then pays the local network 4 cents for two-way call termination. The long-distance operator is effectively able to obtain call origination at the price of 5 – 4 = 1 cent per minute. It is true that the end-user must pay the additional cost of a local call, but is more than compensated by the lower charges for the long-distance transportation component. See Figure 14.

Figure 14. **If two-way call origination is cheaper a long-distance operator may have an incentive to route its calls through a local operator**



Source: OECD.

This possibility for arbitrage between call origination and termination is particularly important when local calling is free. In this case, if there is *any* positive two-way call termination fee the long-distance operator, by interconnecting as a local operator, is effectively able to obtain call origination services at a *negative* price (*i.e.*, the interconnecting operator is paid for call origination rather than having to pay for this service).

Suppose that a long-distance firm needs access to a local loop operator to originate one million minutes of long-distance calls. The one-way call origination charge is, say, 4 cents, and the two-way call termination charge (*i.e.*, the charge at which calls are exchanged between local networks is, say, 2 cents). Local calls are free (*i.e.*, unmetered). In this case if the long-distance operator connects in the conventional way it must pay $400 000 in call origination fees. However, if it connects through a (subsidiary) local exchange company, the end-users will pay for an additional local call (which is free) and the local loop operator will pay the long-distance operator $200 000 for terminating the call. The difference between these two approaches is $600 000.

Whenever the two-way call origination price (*i.e.*, the retail price for a local call less the two-way termination price) is lower than the one-way call origination price, operators have an incentive to arbitrage these prices by re-routing or re-filing one-way call origination traffic through a local network, effectively paying the two-way origination charge.

In the US, the regulator was forced to address exactly this possibility when long-distance operators began offering long-distance connections via a local operator. Customers had to first dial a local number to obtain access to the IXC but, doing so, effectively evaded the one-way call origination charges. The FCC soon blocked this practice. The FCC ruled that local calls which are used to connect to inter-exchange carriers will incur interstate access charges.[4] Another example arises in Canada where, mobile operators, which usually pay a per minute call origination fee to the ILECs, can, by obtaining CLEC status, interconnect with the ILEC on the basis of bill-and-keep, essentially obtaining call origination for free.[5]

Links between two-way call origination and termination

We have discussed the problems that can arise when the regulator seeks to distinguish one-way and two-way traffic. As we will see shortly, only a minority of OECD countries attempt to enforce this distinction in practice. But, even in those countries which do not distinguish between one-way and two-way traffic for interconnection purposes, there remains the possibility of a significant distortion to competition when the tariff for two-way call origination and two-way termination are not the same.

The previous chapter set out several models of two-way interconnection of networks. One of the key features of those models was the assumption that individual subscribers called all other subscribers with equal probability. As a result, as long as the per-call charges on each network are the same, the calls between any two networks are always in balance, *regardless of the sizes of the two networks*. When this assumption is combined with the assumption of reciprocity (*i.e.*, that the charges for two-way call termination are the same in each direction), net termination revenue is always zero in equilibrium.

However, in practice, not all end-users of networks have balanced calling patterns. Some end-users generate significantly more calls than they receive. Others receive significantly more calls than they generate. When some end-users create an imbalance in calls, differences in the structure and level of call origination and termination fees can distort competition for these end-users, in a way that is analogous to the problems that were discussed in the theory of one-way access in the previous chapter.[6]

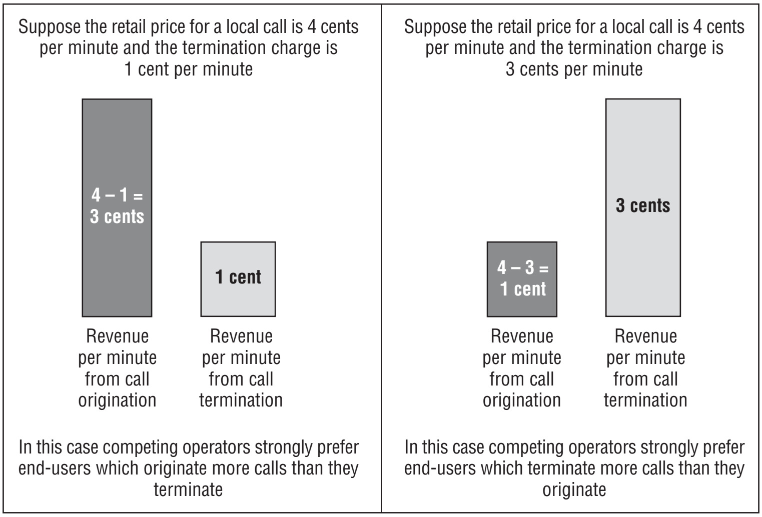
Suppose, then, that different customers have different numbers of outgoing and incoming cross-network calls. Putting aside the costs of providing telecommunications service (*i.e.*, assuming these costs are fixed and independent of traffic), the revenue from two-way call origination is, as discussed before, the retail price ("*P*") for a call less the termination charge ("*T*"). On the other hand, since the termination charges are assumed to be reciprocal, the revenue from two-way call termination is the same termination charge, *T*. The total revenue from a customer therefore depends on the relative proportion of call origination and termination services for that customer and the relationship between the outgoing call revenue $P - T$ and the incoming call revenue *T*.

Note that the outgoing call revenue $P - T$ is larger than the incoming call revenue *T* if and only if the retail price *P* is larger than twice the termination charge *T*. So, in the case when the termination charge *T* is less than half the retail charge *P*, and assuming both charges have the same structure (*i.e.*, both are per call charges, or per minute charges) then it is clear that the carriers will prefer customers who originate more calls than they terminate. (Since $P - T$ is larger than *T*.) In this case network operators will actively seek call producers such as tele-marketers or private pay-phone companies and will tend to eschew call sinks, such as paging networks or Internet service providers.

In the case when the termination charge *T* is more than half the retail charge *P*, the reverse is true – carriers will prefer customers which terminate more calls than they originate (since $P - T$ is smaller than *T*). Network operators will eschew call producers such as tele-marketers and will tend to actively pursue call sinks, such as Internet service providers.

This effect is particularly strong in those countries with free local calling. In those countries *P* is zero so the use of *any* positive metered call termination charges *T* always creates a strong incentive for networks to eschew subscribers which originate more or longer calls than they terminate and to actively recruit subscribers which terminate more or longer calls than they originate.[7]

Figure 15. **The effect of call origination and termination charges on incentives for targeting end-users**



Source: OECD.

In the US, which has unmetered local calling, this problem has become critical. This is described further in Box 6. The FCC writes:

> "We believe that this situation is particularly acute in the case of carriers delivering traffic to ISPs because these customers generate extremely high traffic volumes that are entirely one-directional. Indeed, the weight of the evidence in the current record indicates precisely the types of market distortions identified above are taking place with respect to this traffic. For example, comments in the record indicate that competitive local exchange carriers (CLECs), on average, *terminate eighteen times more traffic than they originate*, resulting in annual CLEC reciprocal compensation billings of approximately two billion dollars, ninety per cent of which is for ISP-bound traffic. Moreover, the traffic imbalances for some competitive carriers are in fact much greater, with several carriers terminating more than forty times more traffic than they originate. There is nothing inherently wrong with carriers having substantial traffic imbalances arising from a business decision to target specific types of customers. In this case, however, we believe that such decisions are driven by regulatory opportunities that disconnect costs from end-user market decisions. Thus, under the current carrier-to carrier recovery mechanism, it is conceivable that a carrier could serve an ISP free of charge and recover all of its costs from originating carriers. This result distorts competition by subsidising one type of service at the expense of others."[8]

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Box 6. **Competition for call sinks and Internet service providers**

This box describes the experience of the US and the Netherlands in competition for Internet service providers that arose from a difference in the call origination and termination charges. The same problem arose in each country, but for different reasons, and with different solutions.

Te US FCC writes:* In most states, reciprocal compensation governs the exchange of ISP bound traffic between local carriers. [Under this regime] … carriers have the incentive to seek out customers, including but not limited to ISPs, with high volumes of incoming traffic that will generate high reciprocal compensation payments…

The resulting market distortions are most apparent in the case of ISP bound traffic due primarily to the one way nature of this traffic, and to the tremendous growth in dial up Internet access since passage of the 1996 Act. Competitive carriers, regardless of the nature of their customer base, exchange traffic with the incumbent LECs at rates based on the incumbents' costs. To the extent the traffic exchange is roughly balanced, as is typically the case when LECs exchange voice traffic, it matters little if rates reflect costs because payments in one direction are largely offset by payments in the other direction. The rapid growth in dial up Internet use, however, created the opportunity to serve customers with large volumes of exclusively *incoming* traffic. And, for the reasons discussed above, the reciprocal compensation regime created an incentive to target those customers with little regard to the costs of serving them – because a carrier would be able to collect some or all of those costs from *other* carriers that would themselves be unable to flow these costs through to their own customers in a cost causative manner.

The record is replete with evidence that reciprocal compensation provides enormous incentive for CLECs to target ISP customers. The four largest ILECs indicate that CLECs, on average, terminate eighteen times more traffic than they originate, resulting in annual CLEC reciprocal compensation billings of approximately two billion dollars, ninety per cent of which is for ISP bound traffic. Verizon states that it sends CLECs, on average, twenty one times more traffic than it receives, and some CLECs receive more than forty times more traffic than they originate. Although there may be sound business reasons for a CLEC's decision to serve a particular niche market, the record strongly suggests that CLECs target ISPs in large part because of the availability of reciprocal compensation payments. Indeed, some ISPs even seek to become CLECs in order to share in the reciprocal compensation windfall, and, for a small number of entities, this revenue stream provided an inducement to fraudulent schemes to generate dial up minutes.

* This material is taken from FCC (2001b), pages 32-34.

Box 6. **Competition for call sinks and Internet service providers** *(cont.)*

For these reasons, we believe that the application [of] ... reciprocal compensation, to ISP bound traffic undermines the operation of competitive markets… We believe that a compensation regime, such as bill and keep, that requires carriers to recover more of their costs from end users may avoid these problems.

The discussion in this report highlights two potential pitfalls with this approach. First, since business customers pay metered rates for local calling in the US, the introduction of bill and keep for local call termination could create the opposite competition distortion – giving rise to strong incentives to compete for businesses which originate more traffic than they terminate – leading to significant traffic imbalances in the other direction. Second, by increasing the difference between one way and two way termination charges, the introduction of bill and keep will increase the incentives on carriers to re route or re file their traffic, potentially demanding greater regulatory surveillance.

In the Netherlands a similar problem arose but for a different reason. In the early days of liberalisation in the Netherlands competing networks signed reciprocal compensation arrangements with the incumbent KPN. The agreed reciprocal termination charges were high relative to the retail price for a local call so that, in effect, the two way call origination charge was less than the one way call origination charge. As in other countries, this gave rise to a substantial incentive for competing operators to attract the business of ISPs. In this competition, competing operators passed on the benefits of the cheaper call origination charges in the form of "kickbacks" to the ISPs.

Subsequently the regulator OPTA decided that KPN's tariffs were not cost oriented and mandated that KPN should lower its origination and termination charges. This reduced the advantage of competing operators in competing for the business of ISPs, but did not eliminate it entirely. There was no competitive pressure on the competing network operators to lower their own termination charges (due to the "terminating network monopoly" problem discussed later in this report) so, in effect, the two way call origination charge remained lower than the one way call origination charge.

KPN complained to OPTA, asking that OPTA impose a reciprocity requirement on competing network operators (this would lower their termination charges and reduce or eliminate this competitive distortion). OPTA ruled however that it would not require reciprocity, but it would allow KPN to raise its retail tariff for calls to rival networks with higher termination charges. KPN did so, effectively raising the two way call origination charge and eliminating its disadvantage in the market for ISP traffic. The rival network operators responded by lowering their own termination charges to match those of KPN.

In effect, competition between ILECs and CLECs is somewhat analogous to the "I cut, you choose" method for dividing a cake. The level of the two-way termination charge determines how the total "cake" (the local call revenue) is divided between the originating and terminating operators. Once this termination charge is set, CLECs can decide which customers they target. They will naturally choose customers that bring them a bigger slice of the cake. If the termination charge is higher than the origination charge they will choose customers who terminate more calls than they originate and *vice versa*.

In summary, the structure and level of two-way termination charges relative to retail charges has little or no effect on the level of competition for the business of those end-users which induce a balanced number of incoming and outgoing cross-network calls. However, in the case of those end-users which induce an imbalance in the number of incoming and outgoing calls (Internet service providers are a prime example) the structure and level of two-way termination charges has important implications for competition. In particular, when local calls have a low price relative to the termination charges, rival local carriers have an advantage in competing for the business of call sinks.

This problem of competitive distortion can be addressed only by correctly adjusting the relative structure and level of call termination and retail charges, especially for those end-users that generate imbalances in call flows. The termination charge should have the same structure and should be exactly half of the retail charge (ignoring origination and termination costs) at least for those end-users that generate imbalances in call flows.[9]

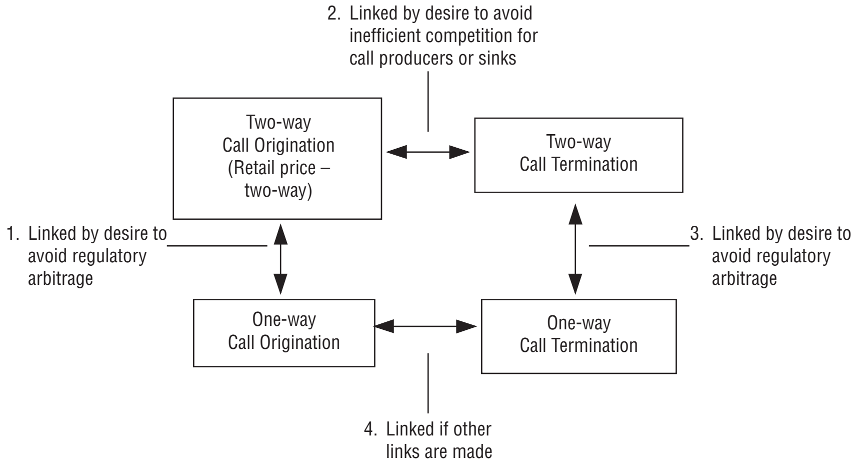
Where the retail charge is too low relative to the termination charge, the problem of competitive distortion can be resolved either by raising the retail charge or lowering the termination charge. As explained in Box 6, the US, which has unmetered local calling, has chosen to resolve the problem by lowering the termination charge. The Netherlands, on the other hand, chose to resolve the problem by raising the retail charge.

The "chain of access charges"

This section has emphasised a number of links between different access call origination and termination charges. We are now in a position to put all these links together to form a chain which we will call the "chain of access charges".

This chain is illustrated in Figure 16. First, two-way call origination and one-way call origination may be equalised in the desire to avoid regulatory arbitrage for call origination (see (1) in Figure 16). Second, two-way call origination and termination may be equalised in the desire to avoid the distortion of competition for end-users with imbalances in outgoing and incoming calls (see (2) in

Figure 16. **"The chain of access charges"**



Source: OECD.

Figure 16). Third, two-way call termination and one-way call termination may be equalised in the desire to avoid regulatory arbitrage for call termination (see (3) in Figure 16). Finally, if all the other links in this chain are set equal, then clearly, one-way call origination and termination must also be equal (see (4) in Figure 16).

If all the links in the chain are taken seriously, *one-way and two-way termination charges and one-way and two-way origination charges should all be equal*. If, in addition, reciprocity applies, all these charges should be equal to half the retail price of a local call.[10] Put another way, if all the links in the chain are important, all the access charges for all the voice services of an end-user are linked to the structure and level of the retail price for a local call.

Some of the implications of the chain of access charges deserve to highlighted:

- First, the chain of access charges is incompatible with an approach under which access charges depend on the distance a call has been transported. Those countries that maintain the chain of access charges must have access charges that are independent of distance. This in turn gives rise to pressure to reduce the dependence of retail prices on distance. In fact, there is strong evidence that retail charges are becoming less dependent on distance. For example, in all the Scandinavian countries (Iceland, Norway, Sweden and Denmark) there is no longer any distinction between local and long-distance calling.

- Second, the chain of access charges is incompatible with a difference in the structure of local and long-distance services. Countries that maintain the chain of access charges must maintain the same structure of charges for all types of calls. If long-distance calls are charged per minute then so must be local calls and *vice versa*. On the other hand, the chain of access charges is compatible with different access charges for different end-users, provided the relationship between access charges and the local retail call price for those end-users is maintained.
- Third, the chain of access charges, combined with reciprocity, implies that local retail charges must be twice the termination charges. *If local calls are unmetered then the one-way and two-way termination charges and one-way call origination charges should all be zero* (*i.e.*, bill and keep for all forms of interconnection). In fact, the US FCC, recognising the problems described above is currently exploring the possibility of extending bill-and-keep for all forms of inter-carrier compensation.[11] As we can see from Figure 17, retail charges for local calls are around twice the termination charges in many EU countries as this theory demands. Retail charges are more than twice the two-way interconnection charge in Germany and the UK and significantly less than twice in Japan. In Germany there is not yet competition for local calls (this will be introduced in 2002). It would be interesting to determine whether there is intense competition for those end-users that originate more calls than they terminate in the UK and more intense competition for call sinks in Japan.

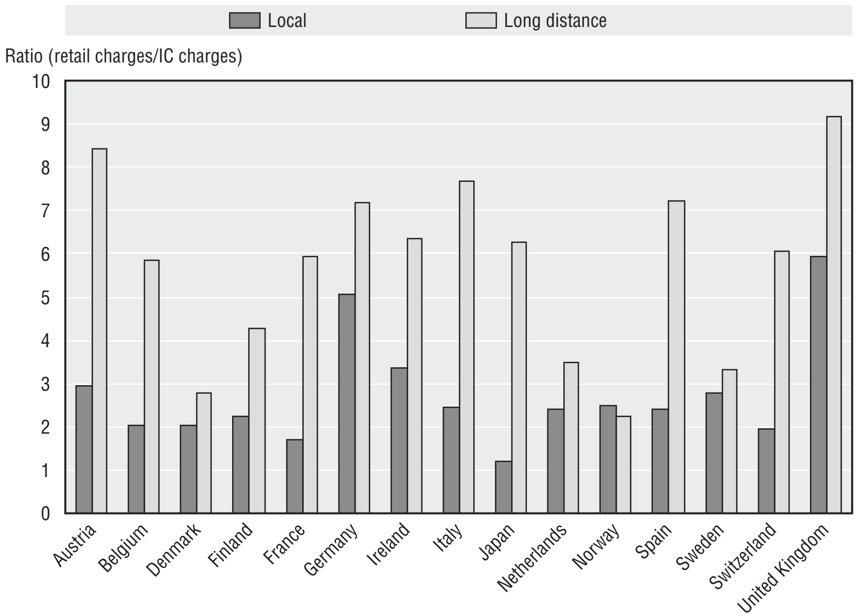
2.1. Broad approaches to call origination and termination prices

Having examined reasons why countries would link the different theoretical access prices, let's turn now to look at what countries do in practice. It turns out that we can group countries into two broad categories: those which maintain all the links in the chain of access prices and those which distinguish between local and long-distance access charges.

Countries which maintain the chain of access charges

A large group of countries choose to maintain all the links in the chain of access prices. These countries set a single access charge, whether for call origination and termination for one-way services (such as long-distance services) or for call termination for the two-way interconnection of networks. These access charges may be differentiated according to the time of day (peak, off-peak), type of user (business, residential) or the number of network elements used, but they do not differentiate according to the category of access service (call origination, call termination, one-way or two-way).

Figure 17. **Comparison of ratio between retail prices and interconnection charges**



1. The local interconnection charges are a weighted average of using 80% of the 5-kilometre charge and 20% of the 20-kilometre charge.
2. The long distance interconnection charges are a weighted average, using 10% of the 5-kilometre charge, 30% of the 20-kilometre charge and 60% of the 50% kilometre charge.
3. For the long distance retail charges, using the tariff that applies to a 100-kilometre call as a proxy for long distance.
4. For the retail charges, using the time of day call profiles to average the charges into peak and off peak.
5. For the retail charges, including call set-up charges assuming call duration of 2.5 minutes but no minimum call charges.
6. For the retail charges, excluding tax charges.

Source: Ovum "Interconnect: Quarterly Update October 1999" and OECD (2001a).

Most European countries fall into this category, including the eastern European countries of Czech Republic, Poland and Hungary.[12] All of these countries have metered charges for local calls. In most cases, as we observed in Figure 17, the retail price for a local call is around twice the access charge.

The principal drawback with this approach is that is significantly limits the discretion of the regulator in the setting of individual access charges. The structure and level of access charges is forced to be the same across a wide range of access services. Theory suggests that if two different retail services have different elasticities of demand they should have different access charges. But this is ruled out if the chain of access charges is maintained. In addition, theory suggests that competition forces the structure and level of

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

retail prices to reflect the structure and level of access prices. The chain of access charges therefore forces many different retail charges to have the same structure. Maintaining the chain of access charges therefore risks either introducing a distortion to competition or limiting the scope for innovative or efficient new retail tariffs.

In fact, as we will see in subsequent sections of this paper, those countries that have adopted this approach have faced interesting competition issues when innovative new tariffs have been introduced. The policy responses have either been to introduce new specific access charges, to tolerate inefficient tariff structures, to tolerate inefficient competition, or, in some cases, to prevent competition entirely.

In its submission for this study, the Netherlands recognises the straitjacket imposed by the chain of access charges and argues for the need to set different charges for (one-way) call origination and termination. The Netherlands writes:

> "The Commission has reached the conclusion that a differentiated approach should be adopted for the tariff regulation of terminating access and originating access. … [T]his differentiated approach takes into account the special nature of terminating access, while at the same time creating the possibility for tariff regulation to be influenced by the effect of actual market conditions. … In the past, the tariffs for both terminating and originating services were set based on the embedded direct cost approach used at the time. The tariff difference between both types of services was exclusively due to the fact that KPN incurred direct costs for originating services that were only relevant to those services. No consideration was given to the specific nature of the terminating service compared to the originating services, nor the possible effects of actual developments in the market, on the way such tariff regulation could best be carried out. The differentiated approach was implemented for the first time in OPTA's decision for KPN's wholesale tariffs for the period 1 July 2001 to 1 July 2002."[13]

Countries which differentiate local and long-distance access

The second broad category which we can distinguish, groups those countries that differentiate between local and long-distance services for the purposes of access charges. The countries in this group include Mexico,[14] New Zealand, Australia, the US and Canada. All these countries have unmetered or free local calling. As noted earlier, if all the links in the chain of access charges were maintained, it would be necessary to set unmetered or zero call origination and termination charges, at least for those end-users which enjoy free local calling.

Rather than set all access charges to zero, these countries have chosen to make a distinction between one-way and two-way call termination (breaking links (1) and (3) in Figure 16). This introduces more flexibility for the regulator, at the cost of having to make arbitrary judgements as to what traffic qualifies as one-way or two-way and at the cost of having to enforce these judgements. The effects of these decisions have been emphasised earlier. As noted earlier, in these countries IXCs and mobile operators have sought to route their origination traffic through CLECs to lower origination charges. In addition, there have been complaints from IXCs about unfair competition from IP telephony traffic, which is allowed to connect at the lower termination rate for CLECs. Finally, the incompatibility of free local calling and per minute termination charges has lead to a substantial distortion in competition for Internet service providers in the US and New Zealand, and regulatory intervention in the US and Australia.

2.2. Further notes on the structure of call origination and termination charges

Having introduced the notion of the "chain of access charges", we are now in a position to examine countries' approaches to regulating specific access services in more detail. Before we turn to this material, however, we first note some related policies that affect the regulation of access charges.

Reciprocity

The theory in the previous chapter highlighted the fact that a requirement for reciprocity in termination charges can play a role in aligning the incentives of networks to agree on a common termination charge. Reciprocity also controls the "terminating network monopoly" problem discussed further below.

Some countries do impose a reciprocity requirement. The US Telecommunications Act imposes on all local exchange carriers the duty to "establish reciprocal compensation arrangements for the transport and termination of telecommunications".[15]

Most other countries do not have a rule imposing reciprocal charges. The Netherlands writes that:

> "The obligation to charge cost oriented terminating tariffs only applies to the regulated firm (KPN) and not to others. Those others are free to decide on the price level for terminating services within 'the boundaries of reasonableness'. … Non-regulated firms can hardly be forced to set the same charge for their terminating services as the regulated firm has to for his. The regulator has not set any rules that require that the terminating access prices should be the same for traffic in each direction."[16]

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Nevertheless, most countries note that in practice agreements for two-way interconnection of fixed-network traffic are almost always at reciprocal rates. It is interesting to ask why this is the case. If the incumbent is allowed to charge different prices for calls to different networks, a possible explanation is that the competing networks are concerned that a higher termination charge would be passed on to end-users in the form of a higher retail charge for calls to their networks. Conversely the incumbent would be reluctant to pass on lower termination charges to its end-users on the grounds that it makes calling the entrant's network more attractive. This is discussed further in the section on terminating network monopolies. As noted in Box 6, when the Dutch regulator allowed KPN to differentiate its retail tariffs according to the destination of the final network, competing network operators chose to lower their termination charges to be in line with those of KPN. The European Commission has recently expressed its concern that termination charges for new entrants are often set on the basis of reciprocity.[17]

Access price-caps

The theory set out in the previous chapter highlighted how the problem of setting the prices of a natural monopoly can be simplified and made more efficient by delegating the task of setting the individual prices to a regulated firm, subject to a cap on an overall basket of prices. Allowing the regulated firm some discretion in the setting of its prices allows it to make use of any private information that it has about demand in order to set prices more efficiently.

Although no OECD country has yet chosen the global price cap approach, almost all OECD countries make some use of price caps in the regulation of end-user prices. On the other hand, the use of price-caps to control access prices remains, at this time, very rare. At the time of writing, only the UK (Oftel) systematically uses price caps as a tool for regulating access prices. These caps are of the common "RPI-X" form. For example, the price cap on call termination charges is RPI-10%. The price-cap on flat rate Internet access call origination is RPI-7.5%.

Capacity-based pricing

As discussed in the previous chapter, in industries where the primary driver of the cost of a network is the network's capacity to carry traffic at peak times, it often makes sense to base the access prices on the capacity allowed to competitors and not on usage. Capacity-based pricing is a form of multi-part tariff, with the fixed part of the tariff proportional to the amount of capacity to which the downstream competitor is granted access. In effect, the essential facility is shared between many competing firms without any of those firms having to incur the entire fixed costs of constructing the facility on its own.

In the telecommunications industry, network capacity is a key driver of costs over large segments of the network. Capacity-based charging, if it can be made operational, therefore seems a logical and efficient approach to pricing interconnection. Throughout this report we have emphasised how the desire to promote competition pushes the structure of retail tariffs to reflect access prices and *vice versa* . One of the primary benefits of capacity-based charging is that it allows the competing firms to break away from the tariff structure of the incumbent.

Capacity-based charging systems have occasionally been promoted in the telecommunications industry. Mercury, a new entrant in the UK, actively promoted discussion of capacity-based charging in the mid 1990s (see Box 7). Touche Ross (1994) reports that the interconnection arrangements between Teleport and Nynex in the New York area at that time took the form of a capacity-based charge.[18]

Capacity-based pricing schemes are becoming more common. As we will see later, many countries have introduced a form of capacity-based pricing for Internet call-origination services.[19] A capacity-based interconnection system has also recently been introduced in Spain, alongside the conventional per minute access charges. Spain notes:

> "Up to August 2001, interconnection prices depended primarily on traffic volume. In August 2001 a new, complementary model of interconnection prices based on capacity was introduced, which will come into effect from 1 November 2001. Operators may adopt one or both systems. This is expected to increase flexibility of operators' final prices, and to promote efficient investment." Under this new capacity-based pricing scheme the regulator "specifies the minimum units of capacity and the minimum length of the interconnection contract, the interconnection services available and the monthly prices per elemental capacity unit. Rights to capacity and minutes of interconnection are tradable."

There may be further movement in this direction in the future. OECD (2001a) notes that "new entrants are arguing that the current per-minute based interconnection charges should be replaced by capacity-based interconnection charges related to the capacity of the facilities used to provide the terminating service. According to the capacity based interconnection charging system, new entrants are charged by their contribution to peak demand in the network which really determines the capacity of networks".[20]

Generic tests against discrimination or price squeeze

Although it is not common in the pricing of call origination and termination services, some countries use a "retail-minus" approach to set access prices. A retail-minus approach sets access prices on the basis of a

Box 7. **Capacity based pricing in the UK: Mercury's proposal**

In the first half of the 1990s Mercury (a new entrant operator in the UK) advocated the use of a capacity based interconnection pricing system. Although this approach was never adopted, many of the arguments of Mercury remain relevant today.

Mercury argued that the interconnection prices should reflect BT's costs and cost structure. Mercury's proposal is to select two cost drivers that account for the largest amount of the total cost incurred by BT in providing a call completion service. … From its own experience, Mercury believes that the two principal cost drivers for all of the network elements that another operator might wish to purchase for call completion are:

- **busy hour erlangs:** drive incremental switching and transmission capacity
- **number of call attempts:** drive signalling, switch processing and billing.

If interconnection payments are to fairly reflect BT's costs then … BT's charges should comprise two elements:

- a **fixed charge** equivalent to leasing "virtual" interconnection capacity based on busy hour erlang forecasts supplied by Mercury (and others)… This would be most fairly collected via either an up front cost (which matches the actual cost of provision plus an acceptable rate of return) or via a recurring rental arrangement;
- a **variable charge** related to the number of call attempts, allowing BT to recover call set up charges…

In this way operators will pay only for those network elements that they cannot provide more efficiently themselves and contribute to BT's costs in the manner in which those costs are incurred. In addition, there are further advantages over and above the existing tariff structure:

- there are incentives for operators to stimulate demand to cover their "fixed" capacity costs. This will permit innovative pricing packages, service packages and customer acquisition programmes;
- the party providing port capacity will have the assurance of revenue from the date the capacity is brought into service;
- each party has a financial incentive to deliver traffic to the closest POI to the destination;
- accurate forecasting is rewarded;
- improved service quality, reliability and ease of network management is assisted;

Box 7. **Capacity based pricing in the UK: Mercury's proposal** *(cont.)*

- the costs of both parties in providing interconnection capacity can be reduced.*

Mercury goes on to say that there should be a punitive (per call) charge for traffic in excess of the purchased busy hour demand (which would apply to both BT and Mercury). It also argues that, since this approach takes some of the risk off BT, BT's cost of capital should be lower.

* Mercury (1993), section 2.

fixed discount off the corresponding retail prices. Under this approach, both the structure and level of retail prices is automatically reflected in access prices.

Even when access prices are not directly linked to retail charges, as in the retail-minus approach, there is often nevertheless, some sort of safeguard or protection against the presence of a "price squeeze". A price squeeze arises when the margin between final prices and access prices is insufficient to allow an efficient downstream operator to compete. A rule against price squeezes can be viewed either as a ceiling on access prices or a floor on final prices, depending on which prices are within the control of the regulated firm.

In Canada, for example, the CRTC has adopted a test known as the "imputation test" to protect against anti-competitive pricing by the incumbent telephone companies. When a telephone company seeks approval of new regulated prices, it must submit an "imputation test" as part of the filing. In essence, this test requires the telephone company to demonstrate the profitability of each of its retail services when accounting for the essential inputs required to provide those services at the same price as those services are sold to rivals.[21]

The Dutch regulator OPTA also uses a "price squeeze" test. "In this test, the retail services offered by the regulated firm are priced as if the regulated firm would have to buy the underlying wholesale services on the same conditions as an interconnecting operator… These wholesale-based retail prices are compared with the actual retail prices the regulated firm uses or proposes. If the former are higher than the latter, a price squeeze occurs – a situation for which a solution has to be found".[22] Using this test it was found that competing operators were not able to provide a competing local retail service on the basis of the wholesale services they had to buy from the incumbent. The Netherlands notes that the regulator "is investigating ways to come to a more structural link between access prices and retail prices, in order to prevent possible anti-competitive effects of the present relation between the regulated firms' retail and wholesale tariff structures".[23]

Ireland also tests for a price squeeze and has addressed complaints in Internet access and retail long-distance calls:

> "Entrants have made repeated allegations that Eircom sets certain of its retail prices and discounts at a level that places a price squeeze on service based competitors. The main services where these questions have been raised are 1891 dial-up Internet access and weekend national calls. In each case the ODTR has established that the price charged by Eircom exceeded the fully allocated historic cost, as represented by the interconnection charge applied to other licensed operators. However, the ODTR has instituted an internal price monitoring process to ensure that any future changes to prices or discounts can be scrutinised promptly."

3. Call origination and termination for voice services

The previous section introduced different types of access services and highlighted certain links between different services. We now turn to look at specific services more closely to explore in more detail the way these access services are regulated.

In this section we focus on call origination and termination for voice services. Because of the links between one-way and two-way services identified in the previous section we will look at these services as a group. In subsequent sections we will look at call origination for Internet service, local loop unbundling and call termination on mobile networks. We start by identifying which call origination and termination services should be singled-out for regulation, and how countries approach this problem in practice. We then look at how these services are priced, comparing the structure of the access prices and the final prices.

3.1. Who is required to provide call origination and termination services?

In theory

What does the theory say about which call origination and termination services should be required to be provided and by whom? This question is more complex than it may at first appear. This complexity is reflected in the variety of practices in OECD countries.

Broadly speaking, call origination and termination should be mandated to and from a given end-user when there is not effective and sustainable competition in the provision of local loop services to and from that end-user. In the case of fixed networks, effective competition in the provision of local loop services is the exception rather than the rule. The scope for competition depends on both the level of demand of each customer and in each geographic area and the economies of scale and scope in serving the geographic area. In most geographic areas the level of demand for all except the largest users is usually insufficient to overcome the substantial economies of scale and scope in the provision of fixed-wire local loops. Even where there is more than one fixed network infrastructure serving a region, usually only the largest users are able to sustain two or more simultaneous fixed connections to different networks. The EC's discussion of the scope for competition in the local loop is set out in Box 9.

On the other hand, in some geographic areas, especially the centre of large cities (known as the central business district or CBD) the level of demand is typically high relative to the economies of scale and scope. Some countries have special regulatory provisions governing local loops in CBDs. Australia, for example, has recently removed certain access obligations in CBDs.[24]

It is important to distinguish between the scope for competition in call origination and in call termination. In the previous chapter we saw that under a particular set of circumstances even very small networks can have an effective monopoly over calls terminating with their subscribers and can have the ability to exploit that monopoly. This problem of "terminating network monopolies" arises when all the following conditions apply:

1. when there is no competition for call termination to a particular subscriber. This is the most common situation since the economies of scale are more important for call termination than for call origination. A new entrant firm might hope to attract a proportion of the call origination business of existing customers away from the incumbent. If the entrant's share of the total call origination business is large enough, it might be able to justify the installation of a competing local loop network. On the other hand, if the new entrant firm has a small proportion of subscribers (say around 1%), it would only hope to capture around 1% of the traffic terminating with any given customer, which is unlikely to be sufficient to justify the installation of a competing local loop network;[25]
2. when the calling party pays the entire cost of the call (*i.e.*, Calling Party Pays) (this is the case for all except free-phone, 0800, or reverse charge calls in most OECD countries);
3. when users care primarily about the price of the calls they originate and not the calls made to them (this implies that "closed user groups" are not significant);
4. it is not possible or desirable to impose reciprocity (the requirement that termination charges are the same on each network); and
5. when retail end-user charges for a call from A to B do not depend directly on the level of the termination charges of the terminating network B.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

When these conditions apply, even very small networks have both the ability and the incentive to raise the charges for termination on their networks and may need to be regulated. This issue arises most clearly in the case of mobile termination in those countries where the calling party pays for calls to mobiles. This issue is therefore discussed more fully below in the section on mobile termination. But the issue also arises in the interconnection of fixed networks. For example, the issue has arisen in the interconnection of IXCs and CLECs in the US.

> "[Another] major issue involves terminating access monopolies. This problem results from the fact that an end-user typically subscribes to only one LEC. Hence, other carriers seeking to deliver calls to that end-user have no choice but to purchase terminating access from the calling party's LEC.[26] These originating carriers generally have little practical means of affecting the called party's choice of access provider. Indeed, … a number of CLECs, whose terminating access charges are not regulated, have taken advantage of this situation by charging terminating access rates that significantly exceed those charged by rate-regulated ILECs… We find that absent intervention, the current disputes between CLECs and IXCs over access rate levels could disrupt the ubiquitous interconnectedness that consumers expect of the public switched telephone network. We adopt, as an interim measure, a detariffing regime in which CLECs may file tariffs establishing access rates only if the rates are at or below a benchmark rate. Rates above the benchmark may not be tariffed. The benchmark is designed to bring CLEC rates closer to ILEC rates over the three-year period that these interim measures are in place."[27]

A similar situation also arose in the Netherlands (described in Box 6). In that case, although interconnection agreements with KPN were initially on the basis of reciprocal charges, when the regulator intervened to lower KPN's termination charges the other networks did not follow suit, highlighting the lack of competitive pressure to lower these charges. As we will see below, many other countries believe it is not necessary to regulate the termination charges on "small" networks (networks which have only a small proportion of connections to end-users in a geographic area).[28]

In practice

How, in practice, do OECD countries designate which services should be provided by which companies? As we will see below, most OECD countries focus primarily on identifying the *companies* that must provide certain services. A minority approach is to focus on the *services* that must be provided.

Box 8. **The scope for competition in local loops**

The EC, in its submission for this report included the following discussion of the scope for competition in local loops:

"The incumbents" local access network is not the only technical infrastructure allowing for the provision of retail services to end users. Other alternatives exist such as fibre optic networks, wireless loops or upgraded cable TV networks. However, none of these alternatives can be considered as an equivalent.

Fibre optic networks are presently only competitive on upstream transmission links and, as concerns the retail distribution network, in special niches like networks connecting office buildings or a narrowly defined geographical area.

Wireless loops appear in the near and medium term future most suitable to address the specific needs of professional clients and small firms, or individual end users with particular needs, but would remain uneconomic for serving the large majority of the residential clientele.

Cable networks which were designed for one way TV need costly upgrades for the provision of two way telecom services, and the provision of high speed services on cable involves customers sharing the capacity of a cable channel, which means that high speed data via cable modems does not offer the same capacity as the copper pair upgraded with DSL technologies, which is dedicated to every single end user. Additionally cable networks do not normally, except maybe in a few countries where the situation would have to be assessed on its own merits, have a nation wide coverage which would allow entrants to serve the same geographic markets as incumbents, whether for traditional voice telephony retail services or new DSL services.

Satellite transmissions are increasingly seen as a possible means to deliver broadband telecommunications services to the end user, but such services would, at least in the short and medium term, be priced at levels that would not make them affordable for the mainstream residential consumer.

Other innovative technologies such as the use of electricity networks could offer challenging alternatives but still have to demonstrate that they are an equivalent and economically viable solution.

While this may over time change, at present none of these alternative networks, nor even their combined use, can for the purposes of delivering retail narrow and large bandwidth telecommunications services be considered as a nation wide alternative to the incumbents' access network.

Given the size of investments required, the absolute cost of nation wide duplication of the incumbents' network with similar population coverage is likely to be a barrier to entry for any competitor. This infrastructure appears

Box 8. **The scope for competition in local loops** *(cont.)*

to be with present technologies economically unfeasible, or unreasonably difficult to duplicate at a nation wide level in a reasonable time period, even for the most important competitors of existing incumbents, in particular incumbent operators from other Member States which develop their activities in neighbouring European countries, alone or in co operation with others. As a result, even though each situation should be assessed case by case on its own merits, most of the time the copper pair is likely to present the features of an essential facility. This situation may change over the long run. New developments shall have to be monitored closely.*

* EC Submission.

In the EU, the approach in the new telecommunications regulatory framework (which comes into effect on 25 July 2003) is to focus the burden of regulation on those companies that are designated as having significant market power in certain markets. Under the new framework, national regulatory authorities must define the relevant markets according to the principles of competition law. (To assist this process, the Commission has issued guidelines on market analysis[29] and has recommended a list of markets which can be subject to regulation.)[30] Where a national regulatory authority determines that a relevant market is not effectively competitive it is required to identify undertakings with significant market power on that market. "Significant market power" is interpreted as equivalent to the competition law concept of dominance. The national regulatory authority can impose additional specific regulatory obligations on companies which are designated as having significant market power (such as requirements for transparency, non-discrimination, accounting separation, obligations of access to, and use of, specific network facilities and price control and cost accounting obligations).[31]

In Japan, access for call origination and termination must be provided by carriers whose total share of the subscriber lines for each municipality or prefecture is over 50%[32] (only NTT East and NTT West qualify). In Korea, interconnection must be provided by carriers whose share in the market is above 50%, whose revenue exceeds an amount set by the regulatory Ministry and which provide "essential facilities for other carriers to deliver telecommunications services".[33]

This approach, which focuses regulation primarily on companies with a dominant position or "significant market power", has the potential to focus attention on the right companies, provided that the market is defined correctly. But the practice of defining the relevant market on the basis of, say, the market share of the firm in the geographic area in which the firm is permitted to operate, or the market share of a firm in a municipality, may be misleading. Since telecommunications services to or from a given subscriber are not a substitute for services to or from another subscriber, the correct market definition is *the market for services to or from a given customer*. A firm may be one of many competing for the business of a specific customer, and therefore may have little or no market power, but may have the vast majority of subscribers in a geographic region. On the other hand, a firm may be the exclusive supplier of local loop services to a given customer, and may have an effective monopoly, even though it has only a few local loops in the region in which it is licensed to operate.

In its recent review of competition in dial-up Internet access,[34] Oftel finds that BT is dominant in the market for call origination due to the fact that it has 81% of residential lines and 86% of business lines in the national market. However the "national market for exchange lines" is almost certainly an incorrect market definition. BT may not be dominant in the provision of local loops to businesses in certain areas and probably has an effective monopoly in certain other areas of the UK.

A firm which is the exclusive supplier of local loop services to a given customer is in a position to deny access by other telecommunications operators to that customer whether or not that firm has a large share of the subscribers in the region in which it is permitted to operate. In the UK, Oftel has designated both BT and Kingston Communications as having SMP. Kingston Communications is a small telephone company that provides services in the Hull region of England. Would the market power of Kingston in the Hull region change simply as a result of a change to a term in its licence that allowed it to operate on a nationwide basis?

Having focused regulatory attention on specific companies, there still remains the question as to what services those companies should be required to provide. Most countries require at a minimum, access to services necessary to provide end-to-end services and any-to-any connectivity. Some countries go further to require access to virtually all the services provided by the incumbent. For example, in Ireland, the incumbent must offer wholesale service parallel to any retail service and, furthermore, must offer any wholesale service for which demand exists.[35] This approach is taken the furthest by the US which requires ILECs to not only provide access to call origination and termination but to provide access to individual network elements of the ILECs' network. In addition, the ILECs have to provide a wholesale service corresponding to all the retail services that they offer.

An alternative approach is that taken by Australia. Australia focuses not on the companies that must provide certain services but the *services* that must be provided (independently of the companies providing those services). In Australia all companies which provide PSTN origination and termination services are subject to regulation for those services. Australia notes: "Generally, the service declarations by the ACCC are not company or operator specific (this contrasts with the policy of dominant carrier regulation prior to 1997, which was focussed on the incumbent Telstra). As such, unless otherwise exempted, the standard access obligations apply to all firms that supply the declared services via their own networks".[36] The risk with this approach is that companies providing services in a genuinely competitive market will be unnecessarily regulated. However, companies which do not wish the access obligations to apply to their services can apply to the ACCC for an exemption.[37]

The US regulatory regime combines elements of both of these approaches (see Box 9). All companies which provide origination or termination *services* are under an obligation to interconnect and, importantly, the FCC has the authority to determine the terms and conditions upon which carriers provide such services.[38] In addition, there are a number of additional regulatory obligations on certain *companies* – those companies the regulatory regime designates as "incumbent local exchange companies" ("ILECs"). These companies are required to provide interconnection, unbundled network elements and access to resale of their services.

Note that national competition authorities may have a role in the decision as to who is required to provide access services to whom. In Mexico, for example, the telecommunications law allows the competition authority to designate which firms in the telecommunications sector are dominant. Once a firm has been designated as being dominant certain additional obligations in the telecommunications law become effective, such as provisions relating to control of prices.[39] In Canada, also, the competition authority has a role in deciding when access obligations should apply – the CRTC has the right to forebear from regulating certain markets when this would be consistent with the overall objectives of the telecommunications law. In developing its criteria for the application of forbearance the "Commission adopted an approach that largely reflected market tests advocated by the Competition Bureau".[40]

3.2. The relationship between access prices and final prices

The theory set out in the previous chapter highlighted the importance of ensuring that the structure of final prices is reflected in the structure of the access prices and *vice versa* . This is illustrated in Figure 18. We look here at the broad features of the structure of final prices and then compare that to the structure of access prices.

Box 9. **Who is required to provide access services to whom in the US?**

In the US, the Telecommunications Act of 1996 ("the Act") distinguishes between 5 principal types of entities: incumbent local exchange carriers ("ILECs"), competing local exchange carriers ("CLECs"); interexchange – including international carriers ("IXCs"); wireless mobile service providers ("CMRSs"); and information service providers, including Internet access providers ("ISPs").

The Act imposes four types of access obligations. The first is the general obligation to interconnect directly or indirectly, to provide end to end services and applies to ILECs, CLECs and CMRS providers. The other three access obligations are imposed specifically on ILECs.

The first obligation on ILECs is the obligation to interconnect with CLECs and CMRS providers to permit each carriers' customers to call one another and to establish "reciprocal compensation arrangements for the transport and termination of telecommunications". Under the FCC's local competition rule making, the access prices for this two way interconnection are set by the states either by using the FCC's forward looking cost methodology or by "bill and keep". In a pending proceeding on intercarrier compensation the FCC is considering whether an alternative system such as "bill and keep" would be preferable to mutual compensation.

The second access obligation on ILECs is the obligation to provide access to "unbundled network elements". These network elements go beyond simple access to the local loop and include access to switch ports and interoffice transport. The network elements that may be leased are determined by the FCC and the states pursuant to section 251(d)(2) of the 1996 Act. Unless the new entrant and the incumbent agree upon the charges for these UNEs, the state public utility commission, as an arbitrator, prescribes the rates in accordance with the statutory procedures set forth in section 252. These rates are to be "nondiscriminatory" and "based on the cost (determined without reference to a rate of return or other rate based proceeding)". The states must follow the FCC's pricing rules in establishing such rates. These pricing rules were affirmed by the US Supreme Court in 2002.

The third obligation on ILECs is the obligation to provide access for the resale of LEC services. Section 251(c)(4) of the 1996 Act requires incumbents to make available at a wholesale discount telecommunications services they provide at retail to users that are not common carriers.

The FCC has also recently imposed certain obligations on CLECs regarding the charges that CLECs can charge the IXCs for originating and terminating their calls. "This proceeding was partly in response to complaints that CLECs were imposing terminating access charges on interexchange carriers that exceeded the rates charged by the incumbent LECs."

Figure 18. **Comparing the structure of access charges and retail charges**

Dependence on:	Structure of retail prices	Structure of access prices
Distance	Is the retail price of a call of 1 000 km more than a call of 10 km?	Is the access price of a call of 1 000 km more than a call of 10 km?
Time	Is the retail price of a call of 10 hours more than the retail price of a call of 10 minutes?	Is the access price of a call of 10 hours more than the access price of a call of 10 minutes?
Time of day	Is the retail price of a call at peak times more than the retail price of a call at off-peak?	Is the access price of a call at peak times more than the access price of a call at off-peak?
Customer	Is the retail price of a call from a business customer more than a call from a residential customer?	Is the access price of a call from a business customer more than a call from a residential customer
Number of calls	Is the price for 1 000 calls of 1 minute different from the price of a call of 1 000 minutes?	Is the access price for 1 000 calls of 1 minute different from the access price of a call of 1 000 minutes?
Calling plan	Is the price of a call of a user on a low-user sheme different from that on a normal scheme?	Is the access price of a call of a user on a low-user scheme different from that on a normal scheme?

Source: OECD.

Final prices

What is the broad structure of end-user or final prices in OECD countries? In all OECD countries, virtually all calling plans for fixed networks feature a two-part structure, with a fixed monthly charge and a usage charge that depends on various components of usage (*i.e.*, a per call fee and/or a per minute fee).

In the case of international and long-distance calls, the usage charge is usually a per-minute charge with no call set-up charge. Although long-distance calls are overwhelmingly priced on a per-minute basis, there is at least some tendency away from a simple linear per minute charge for long-distance calls. For example, many companies have introduced a cap on the price that can be charged for a particular call, or on the total long-distance usage charge per month. For example, Telecom New Zealand has a cap of $2 on the price of national long-distance calls. Under Telecom New Zealand's favourite places scheme, users can pay a fixed monthly charge for unmetered calls to a specific destination. In Canada, BellSouth and Sprint offer a cap of $17 per month on the total national long-distance bill. Telecom Italia offers a tariff option with unmetered national calling. BT, which has metered local calling, offers unmetered local telephony at off-peak times, for a fixed fee.[41]

New tariff packages of this kind continue to be announced all the time. On 16 November 2001, France Telecom announced a plan where for 3 euros per month, users can have unlimited calling to any three designated national fixed-line numbers (including unlimited calling to certain Internet service providers). On 4 December 2001, BT announced that customers would be able to pay a flat monthly fee of almost 30 euros for unlimited free evening and weekend calls to all national and local landline numbers (together with discounts to local, international and fixed-to-mobile calls during the day). In April 2002, WorldCom announced a new service combining unlimited local and long-distance calling in the US for a flat monthly fee of US$50-60 per month.

On top of this trend away from dependence on time, there is an even stronger trend away from dependence on distance. In several OECD countries there is now no distinction between local and long-distance calls – all national calls are the same price – or, put another way, the size of the local calling area is the entire country. This applies in Iceland, Norway, Sweden, Switzerland, Denmark and Ireland and is offered by some new entrants in the UK and Italy.[42] Some operators have taken this principle further and price international calls at the same rate as local calls. For example, one of the new entrants in the UK offers a package in which calls to anywhere in the UK and the USA cost the same – in other words, it costs the same to call next door as it costs to call New York.

In the case of local calls, practices vary. Most OECD countries have a system of minimum fee (or "flag fall") charges combined with per minute charges for local calls. Australia has a pure flag-fall charge – local calls on Telstra's network cost 22 Australian cents, regardless of the length of the call.[43] The remaining countries have free (unmetered) local calls (*i.e.*, no per call or per minute charges – this applies to the US, Canada, Mexico[44] and New Zealand).[45]

In many countries, different end-users face different prices for telecommunications services, either because different prices are offered to different classes of consumers (some telecommunications companies distinguish between business and residential customers) or because the same class of consumers is offered a menu of tariff packages and allowed to choose the one that best fits their needs (for example, many companies offer "low user" options which allow end users to choose tariff packages with a lower fixed fee and higher usage fees).[46] When there are different charges between business and residential customers there may be differences in the monthly rental charge, the usage charges, or both.

Despite the fact that underlying costs vary widely according (amongst other things) to the density of the network, geographic differentiation of retail prices is rare. Canada seems to be the only OECD country to practice systematic geographic de-averaging of retail services. In Canada lines are grouped into 7 bands, according to geographic density. The residential monthly rental for a local loop depends strongly on both the province of Canada in which the subscriber is located and the geographic band of the subscriber's line. See Table A.8 in the appendix to this chapter.

Despite the lack of geographic differentiation, it is common practice to differentiate final prices according to the time of the day or week at which the call is made. There are commonly one, two or three different charging bands, corresponding to peak / off-peak times.

Access prices

Are access prices structured in a way that is consistent with the structure of retail prices just outlined above?

In regard to dependence on distance, we are not aware of any countries where the access charges depend on the distance a call has been carried before being handed off to the terminating network.[47] As long as access charges are not distance dependent, competition provides a strong incentive to eliminate the distance component of end-user charges (since the marginal costs of providing a long-distance call are largely independent of the distance it has been carried). In other words, the absence of distance-dependent access charges could be one of the primary factors behind the significant trend away from distance-dependent retail charges, that was highlighted in section 2 of Chapter 2.

In regard to dependence on time, the majority of long-distance and international calls are still priced to end-users on a per-minute basis. This is consistent with the virtually ubiquitous practice of per minute access charges for long-distance calls. All OECD countries (with the partial exception of Spain)[48] charge per-minute access charges for long-distance calls.

The introduction of capped or unmetered long-distance retail prices breaks this symmetry between the structure of access and retail charges for long-distance calls. This has given rise to competition concerns. For example, in the mid-1990s Telecom New Zealand was charging its rival Clear roughly 2 cents (NZ) per minute for long-distance interconnection per end. Telecom subsequently introduced a new retail tariff which capped off-peak long-distance calls at NZ$ 5, independent of the length of the call. This new tariff structure proved popular with customers but lead to complaints from Clear that for calls longer than roughly 2 hours in duration the interconnection charges would exceed the revenue Clear could obtain from the call.

Similar complaints arose in Australia when Telstra introduced a capped $3 off-peak long-distance calling product. "AAPT alleged that it was providing a competing product to Telstra's $3 capped STD product at a loss because of the structure of Telstra's interconnection prices charged to AAPT. AAPT alleged that, through the capped rate and disparities between peak and off-peak wholesale and retail charges, Telstra was imposing a price squeeze on its competitors, and prevented AAPT from competing effectively in the residential market for long distance national services. The ACCC concluded that Telstra's conduct did not contravene the competition rule."[49]

In regard to dependence on the identity of the end-user, as noted above, a few OECD countries make a distinction between the end-user prices offered to business and residential customers. In addition, many OECD countries offer tariff options that allow end-users to select the basket of tariffs that they want. In contrast, it is rare for call origination and termination charges to depend on the identity of the customer served.[50]

When access prices are not differentiated according to the identity of the end-user, the scope for competition for certain users can be restricted and the ability of the incumbent to offer efficient tariff structures may be impaired. Consider the following example. Suppose that a rival local loop provider wishes to compete with the incumbent by offering its customers a choice of calling plans – one choice, which is targeted at light users features a lower monthly rental fee and a higher per call fee and the other, targeted at heavy users features a higher monthly rental fee and little or no per call fee. Offering a menu of calling options in this way is economically efficient – it could entice low-users onto the network (increasing penetration), while offering more efficient calling plans to heavy users. The key problem is that uniform access charges undermine the ability of the incumbent to raise its usage charges on one group of customers. As Laffont and Tirole note "Uniform access pricing deprives the incumbent of its ability to offer an efficient menu of tariffs tailored to the needs of its clientele".[51]

As discussed in Chapter 1,[52] if long-distance companies are allowed to compete for all customers, whichever calling plan they choose, and if access charges are independent of the calling plan chosen by the subscriber, competition will drive down the price for calls and therefore eliminate the contribution from usage charges towards the fixed costs of serving this customer. Uniform access charges and competition are incompatible with low-user calling plans. If uniform access charges are maintained either the low-user calling plan must be eliminated or competition for these users must be restricted.

Some countries have recognised this problem and have chosen to resolve it, not by allowing access prices to vary with the identity of the end-user, but by restricting competition. The UK, the Netherlands and France have all chosen to restrict competition by preventing customers who chose a "light user" scheme for having a choice of long-distance provider.[53]

Note that it is important to distinguish here between differentiation of access prices according to the identity of the final consumer (or end-user) and differentiation of access prices according to the identity of the access seeker (*i.e.*, the rival or competing firm). In the submissions for this report many countries argued that differentiation of access prices according to the identity

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

of the access seeker raises the potential for anti-competitive discrimination and is explicitly ruled out by laws against discrimination. The Netherlands notes that "the regulated firm is not free to differentiate his termination tariffs depending on his view of possible differences in the business cases of access seekers".[54] Austria argues that two-part access tariffs "would mean that operators would be discriminated according to their interconnection tariff behaviour – smaller operators would possibly pay higher fees than larger ones resulting in high market barriers for new entrants".

These statements are correct if the access prices depend not on the identity of the final customer but on the identity of the wholesale customer (the access seeker). It is at least theoretically possible for the access prices to depend on the identity of the final customer. As long as all companies serving a particular consumer face the same access prices, access prices can differ between consumers while maintaining a level playing field – discrimination is not possible. For example, the access charges for the provision of long-distance service to a customer might consist of a fixed charge of $3 per month and a usage charge of 0.01 cents per minute. An access seeker with 100 such customers would pay a charge of $300 per month. Since these costs increase linearly with the number of customers there is no discrimination against small operators.

In regard to geographic differentiation, as noted earlier, most countries maintain geographically averaged prices. Consistent with this, most countries have chosen geographically averaged call origination and termination charges. An exception is Australia, which differentiates access prices according to geographic location. Call origination and termination charges are divided into four zones – CBD, Metropolitan, Provincial and Rural/Remote.[55] The access charges are lower in the CBD zone than in the Rural/Remote zone reflecting the higher density of the network. This approach reduces the incentive for inefficient network duplication in low-cost areas, but raises the risk that new entrants will target customers in low-cost areas to the exclusion of entry in rural areas.

In regard to differentiation according to the time, most countries differentiate call origination and termination prices in a manner that reflects the peak/off-peak banding of retail prices. The Netherlands notes that the access charge for call termination differentiates between three different time periods which are chosen to be identical to those used by the regulated firm in its retail offer "to prevent price squeeze effects".[56] However, this is not the case in all countries. Australia, Canada and Mexico, for example, make no distinction between peak and off-peak in call origination and termination prices, despite variation in retail prices.[57] In these countries there is a risk that new entry will be focused on end-users which produce most of their calls at peak times.

Let's turn now to look at the scope for competition in the transportation component of local calls[58] – that is, local calls which require both call origination and termination services from the incumbent. Because the transportation component of local calls is limited, not all OECD countries allow or encourage this form of competition. As always, what is important is both the level and structure of access prices relative to retail prices.

As already noted, a number of OECD countries have free local calling. Free local calling is incompatible, on both the level and the structure, with call origination and termination charges that have a per-minute structure. This conflict can be resolved either by adjusting the access charges, the retail charges or by limiting competition. Most of those countries that have free local calls (US, Canada, Mexico and New Zealand) resolve this problem by limiting competition – these countries do not have call-by-call competition for fixed local calls for end-users on free-local-calling plans.[59]

Australia, on the other hand, which has per-call charges for local calls, resolves the trade-off by introducing a special set of access charges for local calls. Specifically, rather than use (cost-plus) per-minute charges for call origination and termination, in the case of local calls, Australia relies on a retail-minus approach which sets access prices on a per call basis. This retail-minus approach maintains a close relationship between final prices and access prices. Since local prices in Australia are not geographically differentiated, are not differentiated according to peak/off-peak but do depend on the identity of the customer (business/residential), the retail-minus approach ensures that this same structure is reflected in the access prices.

Most OECD countries have metered local calls and per-minute access charges, so the structure of the final prices for local calls matches that of the access charges. The primary problem here is not so much the structure of access and retail charges, but the relative level. The problem is that the desire to prevent distorted competition for call producers or call sinks leads the regulator towards setting the call termination charge at around half the retail charge – but this leaves no margin for rival companies to provide local call services. The result is that in all except a few countries, the level of call origination and termination charges has been too high to permit local competition. See Figure 17. A recent OECD report notes:

> "According to OVUM data, it is technically impossible except in the UK and Germany[60] for new entrants to enter the local call market through both ends interconnection with the incumbent local loop. If the [retail price for local calls is less than the sum of call origination and termination charges] a new entrant's marginal revenue does not cover its marginal cost because payments on interconnection charges are greater than revenues to provide local call services".[61]

The next table summarises the structure of access and retail prices for call origination and termination on fixed networks.[62]

Table 2.1. **Call origination and termination on fixed networks**

	Australia		Canada		Czech Republic		France	
Which services offered by which companies must be provided at a regulated price?	All companies providing declared services (whether dominant or not). Declared services include PSTN originating and terminating service and "local carriage service" (resale of local end-to-end service)		Essential/bottleneck services offered by incumbent local exchange companies (ILECs)		In accordance with EC directives, operators with SMP must grant access – SMP is defined as 25% share of telecommunications revenue in a geographic area			
Are the prices for one-way call origination and termination the same?	Yes		Yes		–		Yes	
Are prices for one-way originating and terminating services the same as the price for two-way call termination?	No		Yes (but per trunk for CLECs and per minute for IXCs)		–		Yes	
Are the prices for one-way termination the same as the prices for terminating mobile-to-fixed calls?	No		No (wireless, CLECs and IXCs pay different termination charges)		Yes		Yes	
Is there a requirement for reciprocity in the interconnection of two networks? Is bill-and-keep used?	No		Yes (the termination charge for CLECs and ILECs is generally the same in each direction)					
Is there a requirement that on and off-network calls should have the same price?								
Structure of prices:	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**
Regulated prices can vary within a cap?	No	Yes (CPI-X with sub-caps on Telstra)	Yes, if supported by changes in costs	Yes	No	Yes		
Number of calls (flag-fall or call set-up)?	Yes	Local calls: Yes	No	No	No	No	Yes	
Length of each call (per minute charge)?	Yes (local carriage service: No)	Yes (but some long-distance calls are "capped") Local calls: No	Yes	Yes (long-dist, but caps exist on total bill) Local calls: No (free local calling)	Yes	Yes	Yes	

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	Australia		Canada		Czech Republic		France
Time of day/week (peak/off peak)?	No	Local calls: No Long-dist: Yes	Yes	Yes	Yes	Yes	Yes (three bands)
Geographic location? (*e.g.*, CBD vs rural)?	Yes (local carriage service: no)	No	Yes for unbundled loop service order changes	Yes	No	No	No
Identity of consumer (*e.g.*, business/ residential?)	No (local carriage service: Yes)	Yes (both line rental and call prices	No	Yes (businesses pay more)	No	Yes (rental fees)	No
Is there a relationship between access prices and costs?	Yes: TSLRIC + mark-up + access deficit; Local Carriage Service: No (retail-minus avoidable costs)		Yes: Long-run incremental cost plus 25%		Yes: historic cost (LRIC method being prepared)		Yes: forward-looking average accounting cost
Comments	Given the difference in structure of local calls (unmetered) compared to long-distance calls (metered, with some exceptions), some differentiation in approach to access pricing was necessary, which is provided by the retail-minus access to "Local Carriage Service". The issue of a price squeeze with capped long-distance calls may re-appear.		Canada's approach, like the US distinguishes between different call origination and termination services and requires enforcing arbitrary regulatory distinctions. The introduction of caps on the total long-distance bill may give rise to competition complaints.				

Source: OECD.

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	Germany		Ireland		Italy		Japan	
Which services offered by which companies must be provided at a regulated price?	All public carriers are obliged to respond to interconnection request and access seeker can seek intervention of regulator		Any public telecomms provider designated as having SMP must meet reasonable requests for access at cost orientated rates. Eircom only.		Any operator with SMP must provide call origination and termination on fixed networks, call termination on mobile networks and ULL. (Only TI)		Operator with more than 50% subscriber lines in a region – only NTT East and NTT West	
Are the prices for one-way call origination and termination the same?	Yes		No		Yes		Yes?	
Are prices for one-way originating and terminating services the same as the price for two-way call termination?	Yes		No		Yes		Yes?	
Are the prices for one-way terminating the same as the prices for terminating mobile-to-fixed calls?	Yes		No		Yes		Yes?	
Is there a requirement that on and off-network calls should have the same price?	No							
Is there a requirement for reciprocity in the interconnection of two networks? Is bill-and-keep used?	No		No (but in practice agreements are reciprocal)		No formal requirement but typically agreed			
Structure of prices:	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**
Regulated prices can vary within a cap? Price depends on:	No	Yes	No: RIO	Yes: CPI-8% on PSTN services; In addition there is a sub-cap at CPI+2%	No	Yes (RPI-4.5% until 2002)		
Number of calls (flag-fall or call set-up)?	No	No	No	Yes	No	Yes	Yes	
Length of each call (per minute charge)?	Yes	Yes	Yes	Yes	Yes	Yes (but TI has an unmetered national calling option)	Yes	
Time of day/week (peak/off peak)?	Yes	Yes	Yes	Yes	Yes	Yes	No	
Geographic location? (*e.g.*, CBD vs rural)?	No	No	No	No	No	No	No	

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	Germany		Ireland		Italy		Japan
Identity of consumer (*e.g.*, business/ residential?)	No	Yes	No	No	No	Yes (business customers pay a higher monthly rental)	No?
Is there a relationship between access prices and costs?	Yes: long-run incremental cost plus surcharge for "overhead costs"		Yes: LRIC (based on historic cost)				
Comments							

Source: OECD.

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	Korea		Mexico		Hungary		Netherlands	
Which services offered by which companies must be provided at a regulated price?	Interconnection must be provided by operators with more than 50% market share, revenue in excess of a threshold and which provide "essential facilities"						Operators notified as having SMP – On KPN: origination and termination and local loop unbundling	
Are the prices for one-way call origination and termination the same?	Yes		Yes		Yes (call origination charge is slightly higher due to higher perceived costs)		No	
Are prices for one-way originating and terminating services the same as the price for two-way call termination?	Yes		No		Yes		Yes	
Are the prices for one-way termination the same as prices for terminating mobile-to-fixed calls?	Yes		Yes?		Yes?		Yes (only for KPN)	
Is there a requirement for reciprocity in the interconnection of two networks? Is bill and keep used?	Yes (but bill and keep is not used)		Yes		Yes?		No (but in practice most networks have agreed reciprocal charges)	
Is there a requirement that on and off-network calls should have the same price?	No							
Structure of prices	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**	**Access prices**	**End-user prices**
Regulated prices can vary within a cap?	No		No		No		No	Yes
Number of calls (flag-fall or call set-up)?	No		No		No		Yes	Yes
Length of each call (per minute charge)?	No		Yes		Yes		Yes	Yes
Time of day/week (peak/off peak)?	No		No		Yes	Yes	Yes (same 3 bands)	Yes (3 bands)
Geographic location? (*e.g.*, CBD vs rural)?	No		No		No		No	No
Identity of consumer (*e.g.*, business/ residential?)	No		No		No		No	No; Discounts for large users there are 3 end-user tariff schemes

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	Korea	Mexico	Hungary	Netherlands
Is there a relationship between access prices and costs?	Access prices must follow principles of cost-based pricing			KPN's charges must be cost-oriented
Comments				

Source: OECD.

Table 2.1. **Call origination and termination on fixed networks** *(cont.)*

	United Kingdom		United States	
Which services offered by which companies must be provided at a regulated price?	Focus is on BT and Kingston			
Are the prices for one-way call origination and termination the same?	No (separate price caps)			
Are prices for one-way originating and terminating services the same as the price for two-way call termination?	Yes			
Are the prices for one-way termination the same as the prices for terminating mobile-to-fixed calls?	Yes			
Is there a requirement for reciprocity in the interconnection of two networks? Is bill-and-keep used?				
Is there a requirement that on and off network calls should have the same price?				
Structure of prices	**Access prices**	**Retail prices**	**Access prices**	**Retail prices**
Regulated prices can vary within a cap?	Yes	Yes		
Number of calls (flag-fall or call set-up)?				No (except some business customers)
Length of each call (per minute charge)?				Local calls for most residential users are free (except for local measured service); local calls for business may be metered in some areas; Long distance charges are metered
Time of day/week (peak/off peak)?				
Geographic location? (*e.g.*, CBD vs rural)?				
Identity of consumer (*e.g.*, business/residential?)				
Is there a relationship between access prices and costs?				
Comments	The UK is the only country to systematically use price caps as a tool for controlling access prices.			

Source: OECD.

4. Call origination for Internet services

4.1. Introduction, broad structure of access prices and links with other access services

Let's turn now to look at the pricing of call origination for Internet services. The provision of Internet services requires the combination of (at least) two elements – call origination (over either a low-bandwidth copper wire connection or a high-bandwidth fibre, cable or DSL connection) and an Internet service provider. Of these two elements, Internet service provision is typically competitive while call origination is usually not. We will focus here on narrow-band access to the Internet (via a standard dial-up connection).[63]

Links with other access services

Earlier, in the discussion of the chain of access services, we saw how differences in interconnection charges create incentives for regulatory arbitrage. Internet service providers could theoretically obtain call origination services in at least three different ways. Differences in interconnection charges create strong incentives for Internet service providers to arbitrage across these different approaches.

First, the *ISP may interconnect directly with the incumbent*. As noted in Chapter 1[64] when the ISP connects directly with the local loop operator, the end-user may have a direct relationship with either or both the ISP and the local loop provider. This gives three different possible approaches or "business models": *a)* the end-user may pay for call origination and Internet service provision separately; or *b)* the end-user may pay for the two services simultaneously to the provider of the local loop who then pays a proportion of the revenue to the Internet service provider; or *c)* the end-user may pay for the two services simultaneously to the Internet service provider who then pays a proportion of the revenue to the local loop provider. In countries where local calls are free, the end-user is usually billed directly by the Internet service provider.[65] In countries where local calls are metered,[66] the end-user may be billed by either the local loop provider or the Internet service provider or both.

These approaches are illustrated in the following tables. In each case it is assumed that local loop service costs 5 cents per minute and ISP service costs 10 cents per minute. As this diagram shows, although the payments are different, the three approaches leave all the parties in an identical position and therefore are equivalent. Approach *a)* corresponds to the case where the ISP interconnects with the incumbent local loop operator in the same way as any other (business) user. In this case (under calling party pays) the ISP is neither paid nor is paid for terminating incoming calls.

The end-user cares only about the combined fee for both call origination and Internet service. Provided the total fee is the same, the end-user is indifferent

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Figure 19. **Cash flows under different ISP "business models"**

Approach (a): End-user pays local loop operator and ISP separately for services:

End-user	Incumbent local loop operator	ISP
-5 cents -10 cents	5 cents	10 cents
-15 cents	5 cents	10 cents

Approach (b): End-user pays local loop operator for combined ISP and local loop services:

End-user	Incumbent local loop operator	ISP
-15 cents	15 cents -10 cents	10 cents
-15 cents	5 cents	10 cents

Approach (c): End-user pays ISP for combined ISP and local loop services:

End-user	Incumbent local loop operator	ISP
-15 cents	5 cents	15 cents -5 cents
-15 cents	5 cents	10 cents

Source: OECD.

between these approaches. However, if the Internet service provider is able to choose, it will clearly choose the approach that yields the lowest call origination charge. This charge is the amount paid by the end-user for call origination in approach *a)*, the amount paid by the Internet service provider for call origination in approach *c)*, and the amount retained by the incumbent after paying the Internet service provider in approach *b)*. We will call this amount *O* for call origination charge.

Second, the *ISP may connect with a competing local network, which then interconnects with the incumbent*. In this case, in addition to the approaches outlined above, there is also another possibility – end-users could purchase a local call at the price *P*. The incumbent would then pay the competing local operator a termination charge in the amount *T*. Competition between competing local operators for the business of ISPs forces the local operators to pass this revenue on to the ISPs. So call origination is, in effect, obtained at the price *P-T*. This is illustrated in Figure 20 for the case when a local call costs *P* = 5 cents and the termination charge is *T* = 3 cents. Since the termination charges are passed on to the ISP, the ISP is able to keep a larger share of the total revenue or, equivalently, is able to obtain call origination at a lower price. (See also the lower path in Figure 14.)

Figure 20. **Cash flows when competing local operators can compete to provide Internet access call origination services**

End-user	Incumbent local loop operator	Competing local loop operator	ISP
-5 cents -10 cents	5 cents -3 cents	3 cents -3 cents	3 cents 10 cents
-15 cents	2 cents	0 cent	13 cents

Source: OECD.

Third, in those countries that distinguish between local and long-distance networks, the *ISP could connect to a long-distance carrier* (an IXC) which would then originate calls for the ISP, paying the local network the long-distance call origination rate. In practice, since call origination and termination charges for long-distance calls are usually higher than origination and termination charges for local calls, this last possibility has not been an important factor.[67]

If rival local network operators are to be able to compete on a level playing field for the business of Internet service providers, it must be that the price for purchasing call origination directly is equal to the price for obtaining call origination through a rival local network, *i.e.*, $P - T = O$. or $P = O + T$.[68]

In those countries which maintain all the links in the chain of access charges, and impose reciprocity, this poses no particular problem – the price for origination and termination or ordinary voice calls is exactly half the local call charge, so the Internet service provider is indifferent between connecting directly or connecting through a rival local operator.

For those countries that do not maintain all the links in the chain of access charges (which includes those countries with unmetered local calls), the situation is more complicated. These countries must either restrict competition by preventing ISPs from connecting through competing local carriers, or adjust one of the following prices: the termination charge T paid to rival network operators (at least for calls to ISPs); the origination charge O paid by ISPs to incumbent operators; or the retail call charge P. In Box 6 we discussed the competitive distortion that arose in the Netherlands that was eliminated in part by raising the retail call charge for calls to rival operators.

In the special case in which local calls are free (*i.e.*, P is zero), if there is any positive termination charge T, clearly the origination charge O has to be negative if $P - T = O$ is to be preserved. Put another way, when local calls are free, if there is *any* positive payment for call termination, competing local carriers have a particularly strong incentive to compete for the business of Internet service providers. This distortion to competition can only be avoided if the incumbent local loop operator *pays the ISP* for call origination (rather than the other way around, which would be the normal case). If this is not feasible, this distortion to competition can only be avoided if there is both bill-and-keep for the exchange of local calls *and* call origination for Internet services is free (or unmetered).

We have already raised in this report the problems that have arisen in the US due to competition for ISPs. As set out in Box 6, the combination of unmetered local calling and positive reciprocal termination payments in the US gave CLECs very strong incentives to compete for ISP business and lead to very substantial imbalances in traffic flows between ILECs and CLECs. The FCC's interim solution imposes bill and keep for calls to ISPs. A similar approach was adopted to resolve the same problem in New Zealand. In New Zealand recent interconnection agreements between the main operators distinguish the access charge according to the type of the local call, although call termination charges for regular voice calls is on a (roughly) reciprocal basis, call termination for calls to ISPs is on the basis of bill-and-keep. The New Zealand Commerce Commission writes:

> "Certain types of local calls are categorised as callsinks. Callsinks are basically local telephone numbers that receive many more local call minutes (more than 10 times more) than they originate. Call sinks primarily capture local data calls. No interconnection charges are levied for terminating local calls to call sinks. In other words, local interconnection services for calls to callsinks are provided on a pure bill-and-keep basis."[69]

We will look now at who must provide Internet access call origination services and the relationship between the access and final prices.

4.2. Who is required to provide call origination for Internet services?

In those countries which have free local calling, regulatory intervention to ensure efficient access to Internet service providers has not been necessary – in these countries, Internet service providers can, in effect, obtain call origination for free simply by connecting to the network as ordinary end-users. Indeed, as the arguments above show, Internet service providers have often been able to obtain call origination at a negative price by connecting as an end-user of a competing local network operator.

In those countries with metered local calling, Internet service providers have the option of connecting as normal end-users, but this would normally imply that their customers would pay the full retail rate for a local call in connecting to the ISP. This might place the ISP at a disadvantage compared to the ISP of the incumbent, which might claim that it only has to pay the incumbent for call origination.[70] For this reason, all countries with metered local calling have allowed ISPs to at least only pay the charge for one-way call origination. We are not aware of any cases where a fixed network operator without free local calling is required to provide call origination for other carriers but is not required to provide call origination to Internet service providers or *vice versa* .[71]

Many countries, however, have also introduced special rates for call origination for Internet services. As explained above, this is necessary to prevent distorted competition for the business of Internet service providers when the charge for two-way call origination ($P - T$) is less than the charge for one-way call origination for Internet services (O). In Italy, for example, starting from 2001, the RIO for Telecom Italia will include the price for call origination via 700 numbers which have been dedicated to Internet services and are charged at different prices than the standard retail charges for local calls.[72]

4.3. The relationship between access prices and final prices

How are end-user prices for Internet services structured? The pricing of Internet services to end-users varies, but typically there are a menu of tariff options, with some tariff options based primarily on per-minute charges, and others based primarily on a flat-rate (with intermediate combinations possible).

As we have argued throughout this report, when the services provided by downstream firms compete with the services of an incumbent, access prices must be structured in the same way as retail prices. Per minute charges for call origination for Internet service, for example, are compatible with per minute retail charges for Internet service. Problems have arisen, however, when flat rate retail charges for Internet service have been introduced while retaining per minute call origination charges.

Partly in response to consumer demand, there has been a significant increase in recent years in the number of Internet service providers offering flat-rate packages. At the beginning of 2000, unmetered Internet access was only available in those countries with unmetered local calls (Australia, Canada, Mexico, US and NZ). But by the end of 2000, 11 countries offered unmetered Internet access (including UK, German, Hungary, Korea, Spain and Portugal). Unmetered Internet access is available from some ISPs in other countries.[73]

At the access level, most countries have been slower to introduce unmetered Internet access call origination services. In most countries, call origination for Internet services was, until recently, priced on a per-minute basis. This disparity between the structure of retail prices and the structure of access prices has given rise to competition concerns in several OECD countries.

Box 10 summarises the decision of Oftel in one such case. That case arose from a decision by BT to introduce a flat-rate Internet product called BT Surftime. WorldCom requested a wholesale service from BT to allow it to provide its own comparable flat rate Internet access retail product. WorldCom considered that BT's response was unsatisfactory and complained to Oftel. Oftel determined that a competitor, interconnecting using BT's metered

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Box 10. **Flat rate Internet access call origination service in the UK**

The following text comes from the 26 May 2000 decision of Oftel "Determination of a dispute between BT and MCI WorldCom concerning the provision of a Flat Rate Internet Access Call Origination product (FRIACO)":*

On 7 December 1999 BT announced a new retail product providing unmetered Internet access, called SurfTime, which it would launch in the spring of 2000. The details of the SurfTime product have subsequently been revised, but the underlying principle of an unmetered Internet access product remains. Other operators can, in partnership with BT, provide unmetered Internet access services using BT's SurfTime as the means of conveying calls to the DLE. However, some operators wish to offer a seamless Internet access service to the customer under which the whole of the Internet access product would be provided by the operator. To do this, an operator requires a wholesale call origination product from BT since approximately 80% of all calls made in the United Kingdom originate on BT's network. In order to be able effectively to provide an Internet service to retail customers, OLOs are therefore obliged to interconnect with BT's network. However, such a service is only currently available from BT either on a metered basis, or includes retail elements which are not required by OLOs for them to provide unmetered retail services to consumers;

If OLOs purchase BT's metered wholesale call origination product [and if call volumes increase] … OLOs' payments to BT will increase, but their (flat rate) income from each customer will remain the same. The lack of an unmetered wholesale call origination product therefore has the potential to create a situation under which an operator offering a retail unmetered service would suffer a margin squeeze;

[T]he Director does not accept BT's argument that it faces the same risk as an OLO because its retail business will also be 'purchasing' metered call origination. The Director takes the view that BT does not face similar risks when its position is assessed on an end to end basis. Any losses incurred by BT's retail operation would be purely notional and offset by notional profits in its network business. There are two main reasons for this. First, the metered wholesale charges paid to BT Network (both by BT Retail and by OLOs) are based on *an average* cost. This wholesale charge is substantially above the *marginal* cost incurred by BT's network business in providing the additional call volumes (at any time of day). This means that BT's actual costs do not rise with call volumes in the same way as payments by OLOs using metered call origination;

Secondly, marginal network costs do not grow proportionally with call volumes originating from individual customers. A large increase in call volume from an individual customer is likely to entail greater use of the network's off

Box 10. **Flat rate Internet access call origination service in the UK** *(cont.)*

peak period. During the off peak period, the network has spare capacity. Hence, the marginal cost to BT of a call is at (or not materially different from) zero. An increase in call volume from each individual customer would not necessarily therefore cause BT's costs to rise in line with increased call volumes (or, indeed, at all). However, if OLOs only have a metered call origination service available from BT, they would be required to pay on a pence per minute basis for such extra calls.

For these reasons BT's network business increases its profitability by providing additional volumes and any losses apparently incurred by BT through its retail business as volumes grow and its metered payments to BT's network increase are merely notional and, to a large extent, off set by notional profits enjoyed by BT's network.

The Director believes that there is a serious risk that, without unmetered wholesale interconnection for Internet calls originated on BT's network, competition in the provision of unmetered retail Internet access services will be restricted or distorted.

* Oftel (2000).

Internet access call origination charges would not be in the same position as BT itself (which essentially has access to its own network at marginal cost). In its decision of 26 May 2000, Oftel required BT to provide a flat rate Internet access call origination service to WorldCom.

A similar issue arose in Finland. In Finland, the incumbent operator offered unmetered Internet access at a fixed charge of 125 FIM/month. Rival operators were required to pay call origination fees of 0.03 FIM/minute. An average customer uses 3 hours of access per day, leading to call origination fees of 162 FIM/month. The Finnish Competition Authority is currently in the process of investigating several operators' Internet access tariffs. The EC notes in its recent report on the telecommunications sector in the EU notes that: "flat rate interconnection *must* be offered to new entrants on a non-discriminatory basis by incumbents where they offer their own retail flat rate narrowband Internet access to their customers."[74]

The EC summarises the situation in other EU countries as follows:

> Flat-rate Internet access call origination (or "FRIACO") "is offered by the incumbent at the local and at higher than the local level only in two Member States (Italy, United Kingdom) and is offered at the local level in three further countries (France, the Netherlands, Portugal). In Spain the RIO for 2001

Table 2.2. **Call origination for Internet access**

	USA	Netherlands	Australia
Is call origination for Internet access treated differently from call origination for two-way exchange of local voice calls?	Yes: In order to eliminate regulatory arbitrage, in 2001 the FCC ruled that ISP traffic would be treated differently from other voice traffic through CLECs.	No	Yes: In order to eliminate regulatory arbitrage, in 2001 the ACCC ruled that special tariffs would apply to call termination on ISPs connected to non-dominant (*i.e.*, non-Telstra) networks
Is call origination for Internet access treated differently from one-way call origination (such as for the provision of long-distance calls?)	Yes: the FCC has exempted enhanced service providers from the usage-based charges paid by IXCs. ISPs are instead permitted to pay for their access lines are rates contained in local exchange tariffs.	No (per minute Internet call origination price MIACO for connection at the regional level is the same as carrier selection regional access price)	Yes?
Which firms have an obligation to provide call origination for Internet services?	No need for an obligation – call origination was available until recently at a negative price.	Only KPN	No special designation.
Structure of retail charges	Mostly flat-rate offerings	Both flat rate and per minute offerings	Flat rate per call fee plus generally flat rate ISP charges.
Structure of access charges	Until recently, call origination was available at a negative per-minute rate, but recently the FCC has raised this price to zero.	KPN is obliged to offer both MIACO and FRIACO.	Terminating charges for customer dial-up to ISPs connected to other networks are per minute but subject to a maximum per call
Comments	Under the previous regime, the incentives for CLECs in the US to target Internet service providers was strong.		

Source: OECD.

Table 2.2. **Call origination for Internet access** *(cont.)*

	United Kingdom	Italy	Ireland	Canada
Which firms have an obligation to provide call origination for Internet services?	BT and Kingston	Yes. RIO of 2001 differentiates data calls as 700 number calls. These are 30-35% lower than for voice calls.	Eircom	No special arrangements
Is call origination for Internet access treated differently from call origination for two-way exchange of local voice calls?	No (except for FRIACO)	No	Yes	
Is call origination for Internet access treated differently from one-way call origination (such as for the provision of long-distance calls?)	No	TI	Yes	
Structure of retail charges	Variety of metered and unmetered Internet access plans	Only metered tariff offers are currently available. TI has made a FRIACO offer but this is not yet approved.	Only metered tariffs are available	
Structure of access charges	flat rate Internet call origination rate (FRIACO) has been mandated since June 2000. Metered rate (MIACO) is also available.	Purely per minute charges.	Per minute charges	
Comments	The UK was one of the first to recognise the importance of introducing a flat rate access charge.		Currently developing FRIACO product	

Source: OECD.

introduced a generalised capacity-based interconnection model (applying both to voice and to data), but difficulties have emerged with its implementation. In Germany, the NRA has taken action to impose FRIACO without having so far been able to ensure its availability due to pending court proceedings. No FRIACO offer is available in some countries, despite the fact that the incumbents offer flat rate Internet access to customers (for example in Finland) or flat rate Internet access within certain time periods (for example on Sundays) as part of a bundled offer (Luxembourg)".[75]

Note that flat-rate Internet access call origination (or "FRIACO") can be viewed as a form of capacity-based pricing. Under FRIACO the ISP pays the local loop operator a flat rate which is proportional to the number of simultaneous calls that the ISP can accept. In the Netherlands (KPN) Reference Interconnect Offer in the appendix, the price for FRIACO is € 2 428 per month per 2 Mbit/s link. In addition to the benefits identified above, FRIACO has some of the benefits of capacity-based pricing – in particular, it allows the ISP to set a marginal price close to marginal cost for its own retail service, and it shifts some of the risk of uncertainty of demand for Internet access from the local loop operator to the ISP.[76]

5. Local loop unbundling

Local loop unbundling is the mandating of access to both (physical) ends of the copper-pair local loop on a permanent as opposed to a call-by-call basis. The primary advantage of granting access on a permanent basis is that it makes it technically feasible and financially viable for competitors to install their own equipment on both ends of the line to upgrade the services of the copper-pair. The installation of such equipment would not be viable if it were only used on a temporary or call-by-call basis.

Local loop services can be thought of as comprising two sub-parts – the physical wire or cable connection, on the one hand, and the electronic equipment on each end of the wire, on the other. These are complementary services, one of which is usually competitive and the other of which is usually non-competitive.[77] We therefore have a classic access regulation or essential facility problem.

There are several reasons for mandating local loop unbundling:[78]

1. First, local loop unbundling allows competition to develop in the provision of electronic equipment for upgrading the traditional twisted pair. This is likely to have the effect of increasing the rate of innovation and efficiency in such services. Entirely new technologies for upgrading the twisted pair may come into existence. Local loop unbundling also allows greater competition in the provision of pricing and billing for local loop services, and for combining local loop services with other services.

2. Second, local loop unbundling may accelerate the rate of deployment of broadband services. DSL may be a substitute for some of the existing products of the incumbent operator (such as T1 data lines). If those other products are lucrative, the incumbent operator has little incentive to voluntarily make new investments to increase the rollout of DSL. By mandating local loop unbundling new facilities for delivering broadband services become available and competition in broadband infrastructure is enhanced.
3. Third, local loop unbundling may also act to prevent inefficient investment in one form of bypass of the local loop. If priced correctly, local loop unbundling provides incentive to use existing local loop infrastructure for the provision of broadband services rather than installing completely new duplicative networks.
4. Finally, local loop unbundling also fosters the development of rival networks with direct access to end-users. Such networks have an incentive to compete and to negotiate (two-way) interconnection with each other, without significant regulatory intervention.

As of mid 2001, the vast majority of OECD countries have chosen to mandate local loop unbundling.[79] Local loop unbundling has been an obligation for all member states of the EU since 1 January 2001.[80] OFTEL, the UK regulator, which was initially opposed to local loop unbundling, but which subsequently changed its mind writes:

> "Higher bandwidth access is of fundamental importance to development of new information society services. Technologies such as DSL, cable modems, third generation mobile, broadband fixed radio and digital TV will enable services such as always-on unmetered high speed Internet access, interactive audio-visual services and video-on-demand to be accessible to a wide audience. Enabling these services to develop to their full potential is central to Oftel's primary goal of promoting choice, quality and value for money for consumers. … The best way to achieve the variety of services that consumers want at reasonable prices is to promote effective competition in the provision of access to and delivery of these services… In examining the case for action, Oftel has considered the level of demand in various segments of the market, the supply routes available and whether there are barriers to the competitive delivery of higher bandwidth access and services. The conclusion is that regulatory action is needed to introduce competition into the upgrade of the local loop."[81]

5.1. The relationship between access and final prices

Theory

Which final services use the local loop as an input? There are (at least) two potential end-user services that the rival firms could provide. The first is the

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

standard, narrow bandwidth (mostly voice) telecommunications service. The second is a high-bandwidth DSL telecommunications service.[82] These services are not mutually exclusive. Both are usually provided at the same time.

The theory set out elsewhere in this report makes clear that where the same essential facility is used to provide different services with different elasticities of demand, it is typically efficient for those services to have a different price-cost margin. If a local loop is used to provide both narrow-band voice services and broadband DSL services, the retail or end-user prices for those services should typically have a different price-cost margin. As has been repeatedly emphasised in this report, where the downstream rivals produce a product which is a close substitute for the product of the incumbent, it is important to pay attention to the both the relative level and structure of access prices and the corresponding final prices. If, as just argued, the retail prices reflect a different price-cost margin for different services that make use of the local loop, the access prices should also be differentiated according to the use to which the local loop is put. Failure to differentiate access charges for ULL in this way will lead to a distortion to competition (competitors will naturally focus on the high margin business) and an undermining of the efficient price structure (the efficient price-cost differentials across different services will be eroded).

For this reason, even though this is not the usual practice in OECD countries, the price for access to the local loop should depend, amongst other things, on the use to which the unbundled local loop is put. If the ULL is used to provide narrow bandwidth telecommunications services, it should be priced with reference to the equivalent narrow band service of the incumbent. If the ULL is used to provide a high-bandwidth DSL service, it should be priced with reference to the equivalent DSL service of the incumbent. In the case of line-sharing, where the incumbent continues to provide the low-bandwidth service, the high-bandwidth portion of the ULL should be priced with reference to the difference in the price of the incumbent's bundled high-bandwidth plus voice service and the low-bandwidth voice service alone.

Relatively little information is available at this time about the structure of prices of the incumbent's DSL service in OECD countries. This price is usually geographically averaged, and may differ between business and residential users. More information is available about the structure of retail charges for plain voice telecommunications service – as we have seen above these prices are also usually geographically averaged and sometimes vary between business and residential users.

In addition, it is worth emphasising that it may be important to take into account the fixed/usage structure of the final prices when setting the charge for ULL. The previous chapter emphasised that if the end-user prices of the incumbent (for either voice or DSL services) allow the incumbent to recover a contribution to fixed and common costs from both fixed and usage charges, the unbundled local loop price should also include *both* a fixed and a usage component. The reason for this is as follows. If usage charges are above cost and are making a contribution towards the recovery of fixed and common costs, the loss of revenue to the incumbent from losing a local loop to a rival is not simply the lost fixed subscriber charges but also the lost contributions to fixed costs from the usage charges.[83] An access charge for unbundled local loop which was based on the foregone fixed subscriber charge alone would leave the incumbent undercompensated. Recognising this, the regulator might seek to set the rental price per month for unbundled local loop to include a component reflecting the *average* contribution from usage charges. But, in this circumstance, the rival has an incentive to target only the highest-usage subscribers and to avoid subscribers with lower usage. This simultaneously distorts competition (entry targets only high-usage subscribers) and leaves the incumbent servicing only low-usage customers, which may place the incumbent in a position where it might not be able to recover its fixed costs.[84]

In regard to geographic differentiation of prices, as before, if the scope for competition is to be maximised and if end-user charges are to be preserved, the structure of access charges should reflect the structure of the end-user charges. If end-user prices are geographically averaged, and ULL charges are based on actual costs, the entrants will have a strong incentive to only request unbundled local loops in low-cost areas, intensifying competition in those regions and driving down retail prices in those areas, raising prices in other areas. If geographic de-averaging of end-user prices is the objective (and this cannot be achieved directly by controlling end-user prices), it may in fact make sense to geographically de-average ULL charges. On the other hand, if the regulator wishes to preserve the geographically averaged structure of end-user prices, it is essential to geographically average ULL prices.

Geographic averaging of ULL charges has the disadvantage that it may induce inefficient network duplication in low-cost areas. Entrants will have strong incentives to duplicate existing networks in regions where the incumbent's charges are above cost and little incentive to build duplicate networks (even when it is efficient to do so) in regions where the incumbent's charges are below cost. If it were known for sure what parts of the local loop network are a genuine natural monopoly (for which any duplication would be inefficient) these problems could be resolved through a simple ban on new local loop investment in the natural monopoly areas. But, in practice, it is not possible to determine in advance which parts of the network are a natural monopoly – this depends on demand patterns and technology that are continuously changing. For this reason, regulation of entry is considered undesirable.

 ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

A preferable approach (explained further in the previous chapter) is to set the price for unbundled local loop equal to the "cost" of those loops, and to use taxes on the retail products of the incumbent and its rivals to recover any fixed costs or access deficit. In practice, this would likely imply the establishment of some form of universal service funding mechanism, which "taxed" the revenues of local loop providers in low-cost areas and used those funds to subsidise the activities of local loop providers in high-cost areas.

Practice

When it comes to charging for unbundled local loop there is greater variation in what countries say they do than in what they do in practice. A large group of countries claim that their prices for ULL are "cost based". The EU unbundling recommendation requires that the prices for unbundled access to the local loop shall be "on the basis of cost-orientation".[85]

Consistent with geographically-averaged end-user prices, the regulated tariffs for unbundled local loops are usually geographically averaged (see Table A.7). In fact ULL access prices are usually geographically averaged even in those countries which claim that they are using a "cost-based" or "cost-oriented" approach to the regulation of ULL. The Netherlands, for example, which pursues cost-oriented access prices, unbundles local loop on a geographically averaged basis.

Australia is one of the few countries with geographically-averaged tariffs for end-users, but geographically de-averaged prices for ULL. In Australia, the price of an ULL depends on the line density in different geographic locations. Australia uses four "bands" delineated on the basis of teledensity. These four bands are:

Band 1: CBD areas of Sydney, Brisbane, Adelaide, Melbourne and Perth.

Band 2: Urban areas of capital cities, metropolitan regions and large provincial centres (including other CBD areas not already included in Band 1).

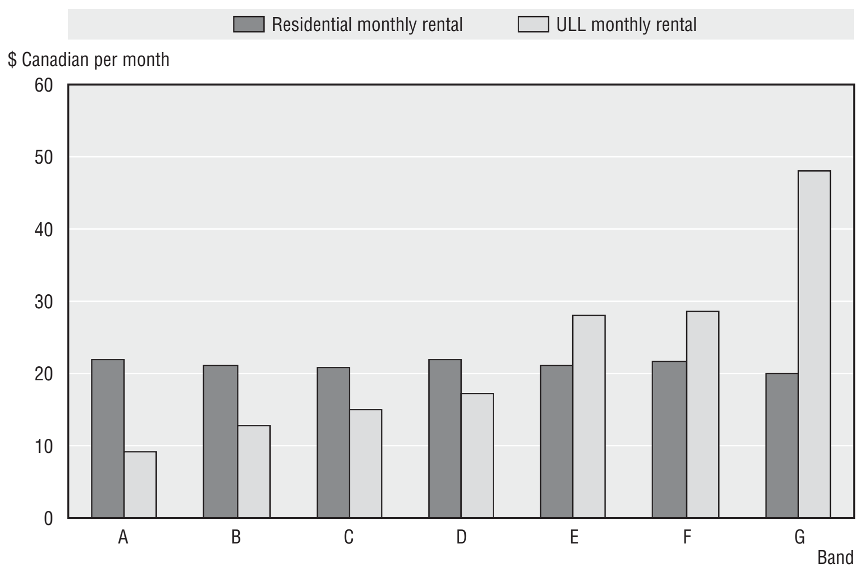
Band 3: Semi-urban areas including outer metropolitan and smaller provincial towns.

Band 4: Rural and remote areas.[86]

This approach has the theoretical advantage of limiting the incentives for inefficient network duplication in low-cost areas. On the other hand, as long as Telstra's final prices remain geographically averaged (and this may be forced to change), de-averaging ULL prices in this way runs the risk that local loop unbundling will be restricted to low-cost areas, such as CBD and urban areas, to the exclusion of customers in rural and remote areas. In fact, the Australian submission acknowledges that "The ULL service is currently used by access seekers to provide high-speed carriage services, such as ADSL, which is largely aimed at the business or corporate customer".

Canada is interesting because it has both geographically de-averaged end-user charges and geographically de-averaged prices for ULL, but the variation in the ULL charges is significantly greater than the variation in the end-user charges. As shown in Figure 21, local loops are grouped into 7 different geographic bands. The access price for ULL exceeds the monthly rental fee for residential local service fee in three of the seven bands. Since the monthly rental for business users is higher, while the ULL price is not, it is not surprising that CLECs in Canada have targeted business customers.[87]

Figure 21. **Residential monthly rental and ULL prices for Bell Canada**
$/month



Source: OECD.

Although retail prices for local loops sometimes distinguish between type of user (business/residential) or even the calling plan of the user, we are not aware of cases where the price for ULL depends on the type of user or the calling plan of the user. Where retail prices are higher for business users, but ULL prices are not, it is clear that rival firms have a strong incentive to target business end-users.

Similarly, despite the theoretical arguments made above, we are not aware of any examples of prices for ULL which include *both* a fixed and a usage component. As already emphasised, as long as call origination and

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Table 2.3. **Local loop unbundling**

	Italy		United Kingdom		Australia		Ireland	
Which firms are required to provide unbundled access to local loops?	Firms notified as having SMP – only Telecom Italia		Firms notified as having SMP – BT and Kingston		Providers of copper local loops (only Telstra)		Firms notified as having SMP – Eircom	
Is access mandated to: fibre-optic loops?	Yes (for a limited timeframe)		Yes		No		No	
ADSL-upgraded copper loops? (*i.e.*, bitstream access with and without sharing)	Yes		Yes		No		Yes	
Structure of prices do prices depend on:	**Access prices**	**Retail prices (for DSL and voice services)**	**Access prices**	**Retail prices (for DSL and voice services)**	**Access prices**	**Retail prices**	**Access prices**	**Retail prices**
Use of the copper (*e.g.*, voice, ISDN, DSL)	Yes (small differentiation according to use)				No	Yes		
Geographic location?	No	No	No	Yes (4 bands)	Yes	No*		
Type of User?	No	Yes (monthly rental higher for businesses for PSTN) No	No?	No	No	Yes		
Choice of calling plan?			No?	No	No	Yes		
Level of Traffic?	No	Yes	No?	No	No	Yes		
Comments				This approach runs the risk that local loop unbundling will be limited to businesses in CBDs		* not mandated by regulation.	No operator has set a retail price as yet. Wholesale price has been set on an interim basis by the ODTR but this is subject to court proceedings.	

Source: OECD.

Table 2.3. **Local loop unbundling** *(cont.)*

	Canada	
Which firms are required to provide unbundled access to local loops?	Major incumbent telephone companies (97-8 local competition decision)	
Is access mandated to: fibre-optic loops?	Yes	
ADSL-upgraded copper loops? (*i.e.*, bitstream access with and without sharing)	Yes (can resell all ILEC services)	
Structure of prices Do prices depend on:	**Access prices**	**Retail prices (for DSL and voice services)**
Use of the copper (*e.g.*, voice, ISDN, DSL)	Yes	
Geographic location?	Yes (7 different geographic bands depending on line density)	Not a one-to-one correspondence
Type of User?	Available to CLECs and DATALECs	Yes
Choice of calling plan?		
Level of Traffic?	No	No
Comments		

Source: OECD.

termination charges are above marginal cost (which is likely to be the case in countries with positive termination charges), there is a strong incentive for rival companies to compete in the provision of local loop services for end-users which are heavier-than-average consumers. It would be interesting to know if this has been the experience of regulators in OECD countries.

Table A.7 sets out the current retail and ULL prices for ADSL and PSTN services in the EU countries. According to this table, certain countries differentiate their ULL prices according to the use made of the ULL. This may be due to differences in the services being provided (the "ADSL" prices listed may include the lease of ADSL modems supplied by the incumbent). From Table A.7 we see that the ULL price is higher than the PSTN price in 6 countries (Germany, Greece, Spain, France, Finland and the UK). In all these countries, the use of ULL to provide PSTN services alone is ruled out. However, in Germany telephone service is provided by competitors on the basis of ISDN. In Sweden, Austria and Denmark, however, the price for ULL is less than the price for current PSTN services.

6. Call termination on mobile networks

Let's turn now to look at the regulation of call termination on mobile networks – especially in those countries with the calling party pays ("CPP") convention for calls to mobile phones.

6.1. Who is required to provide mobile call termination at regulated prices?

Theory

Throughout this report we have emphasised that call termination to a specific end-user is almost always a non-competitive service. In certain circumstances the terminating network, even though it may have only very few end-users, will, if unregulated, exercise its monopoly power over termination to raise the termination charges to monopoly levels. In the US, this issue arises primarily in the interconnection of CLECs and IXCs. In most other countries, the problem arises primarily in the connection of fixed and mobile networks.

As discussed in the previous chapter, in order to have a "terminating network monopoly" problem in fixed-to-mobile calls the following factors must be present:

1. *There must be a bottleneck in termination – i.e.*, call termination on mobile networks must be non-competitive. This is by far the most common case. Although it may be technically feasible for different mobile networks to compete in the termination of calls to a specific mobile handset, this is not currently possible without changing the SIM card of the handset. In addition, consumers must not be able to easily avoid the termination charge on a call from A to B, by, for example reversing the direction of the call (B to A) or calling from another network (C to B). If call termination were competitive and if termination charges are passed on to the calling party, an attempt to raise the termination charge above "cost" would lead the calling party to switch to another terminating firm.
2. *Calling party pays (*"*CPP*"). Under receiving party pays, subscribers to mobile networks pay for both incoming and outgoing calls. As a result, competition between networks will tend to drive down both termination and origination charges. Oftel notes:

   "The overall effect of the CPP principle in the retail market is that, whereas mobile networks have an incentive to keep the price of those services required and paid for by the owner at a level to attract and retail customers, they have less incentive to keep the price of calls to mobiles low. This is because callers cannot take their business elsewhere if dissatisfied as the caller has to use that network to reach that particular number."[88]
3. *Users must care primarily about the price of calls they originate* (and not the price of calls made to them). Even if calling party pays is present, an end-user may care about and be able to influence another end-user's choice of mobile operator. This arises in so-called "closed user groups" where a high proportion of calls are within the group and one member of the group

decides on the choice of mobile operator for all the members of the group. Oftel writes:

"Oftel's evidence suggests that the proportion of residential consumers taking into account incoming call charges when they buy their mobile package is around 13%. … In order to encourage competition between mobile operators in termination charges, the price of incoming calls needs to be an important factor in influencing the mobile owner's choice of network. Research shows that this is clearly not the case, nor do operators suggest that it is. Oftel's research has found that the choice of handset and the price of outgoing services were the two most important factors when choosing a network."[89]

4. *Absence of reciprocity and/or imbalances in call flows between the networks*. As we saw in the previous chapter, reciprocity aligns the incentives of the interconnecting networks and, although not a guarantee of an efficient outcome, may facilitate agreement at rates below monopoly levels.
5. *Absence of direct link between end-user charges and termination charges*. The terminating network monopoly problem is greatly strengthened if the originating network chooses to or is required to reflect only the average termination charges in its end-user origination charges. The FCC writes:

"The terminating access problem is exacerbated by rate averaging policies that are adopted voluntarily by the carrier, or required by regulation such as section 254(g). Rate averaging prevents carriers from passing on termination charges directly to the particular customers whose calls give rise to those charges. Because the originating carrier is effectively unable to pass on termination costs to particular end-user customers or to create incentives for end-users to choose LECs with low termination charges, the end-user who chooses the LEC with the high termination charges does not have an incentive to minimize costs. We note, in this regard, that even if averaging policies were eliminated, it is unclear whether calling parties could, due to transaction cost considerations, effectively induce called parties to choose LECs with low termination charges."[90]

This problem has arisen in the US in the context of interconnection for the termination of long-distance calls.[91] Long-distance calling in the US is primarily calling party pays[92] (unlike calls to mobile phones in the US); IXC-CLEC termination is entirely one-directional, so reciprocity cannot be applied (unlike CLEC-ILEC interconnection); and IXCs are required by section 254 (g) of the US Telecommunications Act to geographically average access charges. So, the conditions for terminating network monopolies are fulfilled.

In the UK, the problem has arisen in the interconnection of fixed and mobile networks. In the UK, fixed-to-mobile calls are calling party pays; it is not possible to impose reciprocity due to the fundamentally different characteristics of the networks; closed user groups are of limited importance and fixed-to-mobile charges are uniform across terminating networks. Thus, the conditions for the terminating network monopoly problem are fulfilled in the UK fixed-to-mobile market.

In the language of competition policy, where there is a terminating network monopoly problem, there is a *separate market for call termination on each mobile operator*. The Netherlands observes that:

> "At the beginning of 2001, KPN Mobile lowered its mobile terminating charges by almost 25%. But the other four mobile network operators did not respond by lowering their prices the same amount. The other operators held their mobile termination charges at their existing high levels. This might indicate that there was not significant switching from the four other mobile operators to KPN. … After three months KPN reversed their action, bringing their mobile termination charges back to their previous high level. This event may be interpreted as a reverse SSNIP – or 'hypothetical monopolist test'. The other four mobile operators were not forced to react to the sharp decline of KPN and were able to maintain their high mobile terminating charges without any significant losses, suggesting that their termination charges were not in the same market."[93]

Practice

Oftel, in its review of the price controls on calls to mobiles concludes that there is a separate market for call termination on each mobile operator and therefore that each mobile operator is dominant in this market and should be regulated. Oftel therefore proposes to regulate the termination charges of all four UK mobile operators (BT Cellnet, One2One, Orange and Vodafone).

In Australia, in December 2000 the ACCC concluded:

> "It is the Commission's view that control over access to GSM termination and consumer ignorance results in mobile carriers sustaining high access prices for GSM termination. The Commission considers that the competitive forces on GSM termination will remain relatively weak, now and in the foreseeable future. The Commission recognises that 'closed' user groups and the possibility of fixed-line callers requesting mobile subscribers to call them back may increasingly place a competitive focus on access prices for GSM termination. However, at this point in time, the Commission considers that the competitive forces on GSM termination are relatively weak."[94]

In Australia and the UK, all the mobile operators have been subject to regulation, even the smallest. In contrast in other countries, the regulators have chosen to regulate only certain mobile operators. For example, in June 1999, OPTA, the telecommunications regulator in the Netherlands observed that the prices for fixed-to-mobile calls were higher in the Netherlands than in other countries and more expensive than mobile-to-fixed calls. As a result it announced that it intended to designate KPN and Libertel as having significant market power, which allows OPTA to control the termination charges for these networks. However, the three other mobile network operators in the Netherlands have been left unregulated.

The EC provides a useful overall summary of the situation in the EU:

> "With regard to mobile termination, regulators have taken a range of measures within the margins set by the current framework to regulate tariffs. In Austria the NRA set mobile termination tariffs on the basis that prices should be 'appropriate', by relying on an imposed cost accounting system. In a number of Member States (the Netherlands, Portugal, United Kingdom), the NRA ordered a reduction in mobile termination tariffs on the basis that it considered those tariffs to be excessive or unreasonable, although it had not designated the mobile operators as having SMP in the national interconnection market. In Finland, mobile operators have not been designated as having SMP in the national market for interconnection, but three of them have been designated with SMP in their own relevant markets, which under national law means that their interconnection charges must be cost-oriented. The cost-orientation of the interconnection charges of two of these operators has been investigated by the regulator. In other Member States the NRA has ordered a reduction which they regard as moving towards the principle of cost orientation, while the mobile operator(s) had been designated as having SMP in the national interconnection market (Belgium, Spain, France, Ireland, Italy, Sweden). In the remaining countries, mobile operators have not been designated with SMP in the national interconnection market and the NRA has not intervened in relation to mobile tariffs."[95]

6.2. Pricing of access to mobile termination

Theory

Theory provides relatively little guidance as to the precise optimal level of mobile termination charges. The previous chapter set out one possible model of mobile competition in which a group of networks competed against each other and simultaneously exchanged essential inputs with a monopoly network. In that model the level of the termination charge directly affected

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

termination revenue on the mobile networks, which in turn affects the level of the fixed charges and therefore mobile penetration. At the same time, the level of the termination charge affects the price of fixed-to-mobile calls and therefore distorts the price of fixed-to-mobile calls relative to fixed-to-fixed calls.[96]

Unfortunately, this model does not provide clear guidance as to the level of the termination charge. The fixed costs of establishing and operating a mobile network are large and are not currently recovered entirely through fixed charges on mobile subscribers. These fixed costs must therefore be recovered through a mark-up on usage – whether mobile-mobile, mobile-fixed or fixed-to-mobile. Without further information on elasticities of demand it is not possible to determine how these fixed costs should be recovered.[97]

Practice

In practice, mobile termination charges are regulated in a number of different ways. The UK's Oftel is of the view that mobile call termination charges should be regulated in the light of the long-run incremental cost of providing the call termination service, but instead Oftel has chosen impose a price cap of RPI-12% each year for the next four years on BT Cellnet, One2One, Orange and Vodafone.

Australia has adopted a unique approach of regulating mobile termination charges on the basis of the movement in a basket of other charges for mobile services. In its July 2001 report on access pricing for GSM termination:

> "the ACCC determined that TSLRIC pricing was not appropriate for the GSM originating and terminating services. … The ACCC determined that a retail benchmarking approach would likely be used in access disputes to determine the access prices for GSM originating and terminating services. This provides that changes in each mobile carrier's access prices will be benchmarked against the retail price movements of its overall mobile package (access and outgoing calls).[98] The initial starting point for the glide path created by this pricing rule will be the lowest current access prices for the GSM originating and terminating services in the market. The approach uses a reasonable proxy for the efficiency improvements and competitive pressures on mobile retail prices to provide a safety net that access price falls continue to occur, particularly for the GSM terminating service. This is likely to reduce the opportunities for anti-competitive pricing and improve allocative efficiency, best promoting the long-term interests of end-users at this stage."[99]

6.3. Further notes on access to mobile services

The interconnection of two mobile operators is a classic two-way access problem. Should the termination charges for mobile-mobile interconnection be regulated? And, if so, should they be the same as the fixed-to-mobile termination charges?

According to the theory in the previous chapter, the common reciprocal termination charge in the interconnection of two competing networks should be set equal to marginal cost. Since the marginal cost of termination in a mobile network is close to zero (at least at off-peak times) this suggests that bill-and-keep might be appropriate for mobile-to-mobile interconnection.

However, as this paper has emphasised in various ways, the feasibility of this approach depends, amongst other things, on whether or not it is possible for the terminating network to distinguish traffic originating on another mobile network or traffic originating on the fixed network. If it were not possible to make such a distinction, mobile operators could pick up calls originating on the fixed network which are destined for subscribers on rival networks, collect the termination charge and then hand off the call to the rival network for termination at no charge. Competition of this form would drive the fixed-to-mobile termination charge down to zero.[100]

In practice, it seems that most countries do not distinguish mobile-to-mobile and fixed-to-mobile call termination. That is, for most countries, the regulated price for terminating a mobile-mobile call is the same as the price for terminating a fixed-mobile call. In the UK, for example, the price cap on mobile termination applies to both fixed-to-mobile and mobile-mobile termination. In contrast, in Korea, up until the end of 2001, the termination charge for a mobile-to-mobile call was the same for all mobile service providers so bill-and-keep was used between mobile service providers. From January 2002 different termination charges were set for the incumbent and the entrants.

In the US, which has receiving party pays ("RPP") for calls to mobile operators, the interconnection of fixed and mobile networks is analytically identical to the interconnection of two fixed or two mobile networks. The US currently applies reciprocal termination charges for the interconnection of fixed and mobile networks and is considering a move to bill and keep.[101]

We have focused on the pricing of call termination on mobile networks. Can we say anything about the pricing of call origination for fixed-to-mobile calls? The price for call origination can be viewed as the difference between the retail charge for a fixed-to-mobile call and the mobile call termination charge. As with ISPs (see the discussion in Chapter 2, section 4.1 above), mobile networks could theoretically obtain call origination by connecting directly with an incumbent and paying for call origination services or by

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Table 2.4. **Call termination on mobile networks**

	Australia	Italy	Ireland	United Kingdom
Which mobile firms must provide call termination at regulated prices?	All GSM mobile network operators (Telstra, SingTel Optus and Vodafone)	Only TIM and Omnitel must provide termination at cost-oriented rates (termination charges on Wind and Blue and UMTS operators are not regulated)	Obligations on Eircell on EsatDigiphone. They are required to have cost-oriented rates	Controls apply to all four operators: BT Cellnet, One2One, Orange and Vodafone
Is fixed-to-mobile call termination priced the same as mobile-to-mobile call termination?	No	Yes – all operators pay the same termination charges	Yes	Yes
Are mobile retail charges regulated?	For Telstra only	No		No
Is discrimination in the price of on and off-network calls allowed?	Yes and such discrimination does occur.			
Structure of call termination charges	Termination charges are regulated on the basis of the movement in the price of a basket of mobile call charges.	Current rate fixed by the regulator and future rates fixed on the basis of current cost model drafted by mobile operators – under review at the moment.	Under review	Price cap: RPI-12 % until March 2006
Do termination charges vary peak/off-peak, per minute or per call?	Per minute charges – no variation peak/off-peak or call set-up charges.	Per minute charges differentiated peak/off-peak.		Per minute charges differentiated peak/off-peak and weekend.
Comments				Oftel was one of the first regulators to recognise the lack of competition in call termination charges and the first to extend regulation of call termination charges to all mobile networks.

Source: OECD.

Table 2.4. **Call termination on mobile networks** *(cont.)*

	Canada	United Kingdom
Which mobile firms must provide call termination at regulated prices?		Controls apply to all four operators: BT Cellnet, One2One, Orange and Vodafone
Is fixed-to-mobile call termination priced the same as mobile-to-mobile call termination?		
Are mobile retail charges regulated?	No (forborne)	
Is discrimination in the price of on and off-network calls allowed?		
Structure of call termination charges		Price cap: RPI-12 % until March 2006
Do termination charges vary peak/off-peak, per minute or per call?		
Comments		Oftel was one of the first regulators to recognise the lack of competition in call termination charges and the first to extend regulation of call termination charges to all mobile networks.

Source: OECD.

connecting to a competing local carrier, in which case the competing local carrier would normally be paid for terminating the call from the incumbent. We mentioned earlier the fact that in Canada mobile operators have the opportunity to connect as CLECs and, if they do so, they are able to obtain call origination at no cost (interconnection of ILECs and CLECs in Canada is on the basis of bill and keep).

In New Zealand, until recently the charge for originating fixed-to-mobile calls was considerably higher than the charge for originating either local or long-distance calls. The New Zealand government decided to intervene by allowing other interconnecting operators to carry fixed-to-mobile call traffic. This will, in effect, allow mobile operators to obtain call origination at either the long-distance rate or the local rate (depending on whether further restrictions are imposed).

7. Conclusions

This chapter has sought to integrate the theory presented in the previous chapter with the practice of access pricing in OECD countries. The theory has highlighted several potential pitfalls for regulators, such as the need to pay attention to both the relative structure and the level of access and final prices, or the possible need to regulate termination on even very small networks.

Over the last five years there has been a notable convergence in OECD regulatory regimes as experience has highlighted weaknesses in the approaches

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

of different countries. This process is surely far from complete. Incumbent operators are continuing to introduce new and innovative tariffs, which are often both attractive to end-users and likely to harm the interests of their rivals. Examples include the new unmetered national calling options being introduced in Europe. Such developments may force further changes in the structure of access prices.

There is a possibility that in the longer term a new approach to access pricing may be needed. If, as this report argues, it is necessary to maintain close attention to the structure of access charges relative to retail charges then either incumbents must be prevented from exercising their discretion to introduce new and innovative tariff structures, or the setting of access charges must be made more flexible. Perhaps even to the point of giving incumbent operators the same flexibility to set access charges that they currently have to set retail charges. It is likely that future developments will be in one of two possible directions – either greater use of price-caps for access charges (perhaps combined with retail charges as in the global price cap) or through greater use of capacity-based charging.

Notes

1. Throughout this report, to keep things simple, we will not distinguish between different types of switches (*e.g.*, local exchange, tandem, etc.). The simple telecommunications networks that we have in mind have just three elements – local loops, local exchanges and connections between exchanges. We will put to one side issues related to the level or layer of the network at which interconnection occurs.
2. In the US, this problem also arises with the termination of IP telephony calls. "The fact that an IXC must pay access charges to the LEC that originates a long-distance call, while an ISP that provides IP telephony does not, gives the provider of IP telephony an artificial cost advantage over providers of traditional long-distance service." See FCC (2001a), page 7.
3. See OECD (2000b), page 53.
4. "Under 'Feature Group A' access the caller first dials a seven-digit number to reach the IXC, then dials a password and the called party's area code and number to complete the call. …[In this case] the service the LEC provides is considered interstate access service, not a separate local call". FCC (2001b), page 29.
5. To date, however, only one mobile operator has chosen this option.
6. Intuitively, this result can be understood as providing the link between one-way and two-way access problems. In the extreme case where an end-user only received calls (an Internet service provider might be a good example), the appropriate theory is clearly the theory of one-way access. In this section we emphasise that some key results of that theory carry over to the case of those subscribers which generate an imbalance between outgoing and incoming calls.
7. One country in its submission to this chapter noted that the percentage of users with imbalanced calling patterns is quite small. The difference of competition

intensity and attractiveness of these customer groups may not be sufficient to push the regulator to adjust the tariffs in the way indicated here. "It does not mean that imbalanced traffic could not cause a competition problem originating in the different levels of the interconnection tariffs but the effect of the imbalanced traffic seems to be negligible compared with other economic issues."

8. FCC (2001b), page 4-5, emphasis added.
9. Incorporating the effects of origination and termination costs is straightforward. Suppose that the marginal cost of call origination is c^O and the marginal cost of call termination is c^T. Then this competitive distortion can only be eliminated when $P - T - c^O = T - c^T$ or when $T = \frac{P - (c^O - c^T)}{2}$ In other words, if origination costs are higher than termination costs, the termination charge should be less than half the retail price (and *vice versa*).
10. Putting aside the marginal costs of call origination and termination.
11. See FCC (2001a). The New Zealand Commerce Commission has proposed the use of bill-and-keep for all local calls. See Commerce Commission (2002).
12. See for example, the access prices of Germany, Table A.4, the Czech Republic, Table A.3 and Poland, Table A.6. Hungary has slightly different charges for call origination and call termination because the government decided that call origination involved a higher cost that call termination – "the Government decided to set the origination price on a higher level than the termination price – this was a move to cover the local access deficit but it was based on the estimation that a call origination has higher cost than a termination as this also includes customer care". Hungary submission.
13. The Netherlands has chosen to set call termination charges on the basis of a forward-looking long-run incremental cost model, while call origination charges are set on the basis of a top-down cost accounting model. The structure of charges for call origination and termination in the Netherlands is set out in Table A.1 in the appendix. As can be seen from Table A.1, the difference in the call origination and termination charges in the Netherlands is, in practice, relatively small.
14. See, for example, the access charges of Mexico in Table A.1.
15. US Telecommunications Act Sec. 251(b)(5).
16. Dutch submission.
17. The European Commission in its 8th Implementation Report (published in December 2002) notes that "Concerns still arise from the fact that in certain circumstances the charges which new entrants can levy for termination on their fixed network are based on reciprocity … despite the fact that those operators are not subject to the obligation of cost orientation and do not necessarily provide a similar interconnection service". EC (2002), page 22.
18. This charge was primarily a fixed charge per month per DS1 (24 voice grade trunks) circuit. This charge was divided according to the volume of traffic in each direction at the busy hour (so, if at the busy hour there was equal traffic in each direction, the charges would net out to zero).
19. See section 3.4.3.
20. OECD (2001a), page 22.
21. Canada submission.

22. Dutch submission.

23. The Spanish submission notes that in Spain "the relationship between access prices and retail prices is always examined, especially when approving discount programmes proposed by the incumbent".

24. The ACCC is considering lifting the obligation to provide access to "local carriage service" (an end-to-end local service) in the CBDs of Sydney, Melbourne, Adelaide, Brisbane and Perth. In Canada, when the CRTC mandated access to local loops in remote and rural areas in 1997, it also concluded that local loops in CBDs did not meet the definition for "essential facilities" but that, nevertheless, access to these local loops should be mandated in the same manner as other local loops, but only for a period of five years. In 2001 the CRTC extended this sunset period for these so-called "near-essential" facilities, without specifying a termination date, until such time as the market for such facilities is sufficiently competitive.

25. This argument in effect assumes that there is no competition for terminating calls. If the calling party could choose the terminating network, the entrant might be able to capture a much larger share of the terminating traffic, perhaps even enough to justify new infrastructure investment.

26. "With regard to wireless networks, we recognize that, where a customer subscribes to both a wireless and a wireline network, the wireline network does not have a complete monopoly over termination. We believe, however, that the customer's possession of a wireless number does not completely resolve all terminating access issues. Since wireless customers are generally charges per-minute rates when they receive calls, they have an incentive to receive calls on their wireline phones. TO encourage this, wireless customers frequently withhold their wireless numbers, both directly, and from directory databases. In turn, many callers respect this preference by choosing to call the customer's wireline number before trying the wireless number."

27. FCC (2001a), page 7-8.

28. Some concern has been expressed about whether or not to mandate access to new, upgraded, facilities provided by incumbent operators, such as fibre-optic cables, cable television networks or DSL services. In Italy, for example, the competition authority became concerned that mandating unbundled access to fibre-optic local loops could undermine the incentives for new entrants to develop their own broadband infrastructure at a local level. The Italian competition authority mandated, instead, access to the civil infrastructure (pipes and ducts) in order to promote the emergence of alternative network infrastructure. This issue is discussed further in Chapter 3.

29. Commission guidelines on market analysis and the assessment of significant market power under the Community regulatory framework for electronic communications networks and services. *Official Journal C 165, 11.07.2002, pages 6-31*.

30. Commission Recommendation of 11.02.2003 On Relevant Product and Service Markets within the electronic communications sector susceptible to *ex ante* regulation in accordance with Directive 2002/21/EC of the European Parliament and of the Council on a common regulatory framework for electronic communication networks and services.

31. See Directive 2002/19/EC of the European Parliament and of the Council of 7 March 2002 on access to, and interconnection of, electronic communications networks and associated facilities.

32. Telecommunications Business Law 38-2, Regulation for Enforcement of the Telecommunications Business Law 23.

33. The Korea submission notes that "the criteria for defining essential facilities needs to be made clearer".

34. Oftel (2001a), page 36.

35. Irish Submission.

36. Australia submission.

37. In Germany, every telecommunications carrier has an obligation to respond to request for interconnection and, in the event of disagreement, can appeal to the regulatory authority. The authority notes, however, that it has never been obliged to order the interconnection of the networks of two non-dominant companies. German submission.

38. For example, as mentioned earlier, in response to complaints that competing local exchange carriers ("CLECs") were imposing access charges that exceed the rates charges by ILECs, the FCC recently extended the regulation of call origination and termination to cover the services provided by CLECs.

39. See chapter 5 in OECD (1999).

40. Anderson *et al.* (1998), page 190.

41. See Chapter 7 in OECD (2001c). Even mobile companies have started offering unmetered mobile calls, especially at off-peak times.

42. See Chapter 7 in OECD (2001c). On 1 May 2002 the Swiss incumbent operator Swisscom eliminated all distance dependent tariffs for national calls and now only offers distance-independent national calling.

43. The prices of local calls on other networks are not restricted to be per call, but the vast majority of local calls are on an untimed basis.

44. In Mexico, the first 100 local calls are free.

45. New Zealand has free local calling for residential (but not business) customers under the standard calling plans.

46. In other words, most operators use some combination of second and third degree price discrimination.

47. It is true, however, that many countries set different termination charges according to the level of the network at which the originating network interconnects. Charges will tend to be higher for connection at a regional level compared to a local level, and higher for connection at a national as opposed to a regional level because more transportation services are required from the terminating carrier. The termination charges themselves do not differ according to the distance the call has been carried *prior* to being handed over. If longer-distance calls were systematically connected at a higher level in the network then there would, in effect, be higher termination charges for longer-distance calls. But there is no reason why this would necessarily be the case – a sufficiently developed network which is able to interconnect at the local level will pay only local termination charges whether the call is an international or a local call.

48. As explained earlier, Spain has a capacity-based system that operates in parallel to a conventional per-minute system.

49. Australian submission.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

50. We will see later that other access services may depend on the identity of the end-user
51. Laffont and Tirole (2000), page 112.
52. See example 4 in Box 4 of Chapter 1.
53. Laffont and Tirole (2000), page 111. To be effective, this restriction would have to apply to rival local loop operators as well as the incumbent.
54. Dutch submission.
55. Australia and Canada also have geographically de-averaged ULL charges, as explained later.
56. Dutch submission. See Tables A.1, Tables A.3 and Tables A.6 in the appendix.
57. There is no differentiation between peak and off-peak in local calling prices in Australia.
58. For the purposes of this paper we will assume that a local call is defined by the boundaries of the local calling area of the local carrier. If the local calling area is small enough to feature only one local exchange there will be no transportation component in local calls, as all local calls originate to and terminate from the same point.
59. Of course, fixed local calls do compete, to some extent with mobile calls. The US notes that while there is no call-by-call competition for local calls in the US (as there is for long-distance service) there is competition for the total calling package (*i.e.*, monthly rental plus usage charges).
60. Despite being technically possible, Germany does not have competition for local calls. This is expected to be introduced in 2002.
61. OECD (2001a), page 23.
62. One country, in its submission on this chapter, notes that it is not "useful or desirable" to introduce a finer price discrimination in access prices than is already present and that the introduction of two-part pricing at the access level would "only lead to discrimination". While it is theoretically possible to carry out two-part pricing in a way which does not lead to discrimination, this may not be feasible in practice. Furthermore, there may be practical difficulties with finer differentiation of access prices which are not discussed here. It may be necessary to trade-off the costs of implementing more sophisticated access pricing schemes with the problems raised in the paragraphs above.
63. In its review of competition in the dial-up Internet market Oftel (the UK telecommunications regulator) distinguishes a third component – call termination to the ISP. In other words, in Oftel's view, Internet service requires a combination of three elements: call origination, call termination and Internet service provision. Oftel (2001a).
64. See section 1.1.
65. Indeed, in these countries, the Internet service provider can interconnect to the incumbent as an other ordinary end-user.
66. Or where access to the ISP is via a metered long-distance call.
67. The FCC did consider whether ISPs should be required to pay long-distance access charges but decided instead to exempt "enhanced service providers from the

usage-based charges paid by IXCs. ISPs are instead permitted to pay for their access lines at rates contained in local exchange tariffs". US submission.

68. More generally, if there is a positive marginal cost of termination then.

69. Commerce Commission (2002), page 14.

70. Or just the marginal cost of call origination.

71. As a consequence, any weaknesses in the definition of which firms have obligations to provide call origination for long-distance services is carried over to Internet services.

72. A related issue, which will not be discussed further in this report, is the issue of access to high-bandwidth local loops for the purpose of call origination for Internet services. Should cable television providers or providers of fibre-optic local loops be required to provide access to third-party ISPs? In Canada, in 1998 the CRTC "determined that CATV operators had to provide open access to their facilities to enable third-party ISPs to provide high-speed cable modem services". See OECD (2001b). "This decision was followed up in 1999 with a requirement that incumbent cable companies offering cable modem services were required to resell these services to ISPs at a CRTC mandated discount of 25%."

73. See Chapter 7 in OECD (2001c).

74. EC (2002), page 21, emphasis added.

75. EC (2002), page 22.

76. A local loop operator must size its network according to expected demand. Under metered access services the revenue depends more closely on demand so the local loop operator faces a greater risk from demand uncertainty – under flat rate access services some of the risk of demand uncertainty is shifted to the interconnecting operator.

77. The US notes that it may not always be possible to have competition in the equipment used for upgrading the local loop, especially when the incumbent locates equipment in remote terminals where it is hard for competitors to co-locate.

78. See OECD (2002).

79. New Zealand is one exception.

80. Regulation 2887/2000 of 18/12/2000 on unbundled access to the local loop.

81. OFTEL (1999), Chapter 2.

82. We are not distinguishing here calls to Internet service providers and other services that might be provided.

83. Although we are arguing here that the ULL price should include both a fixed and a usage component this does not imply that the regulator should use the ECPR or a similar rule which ensures that the incumbent can recover in usage charges *all* of the lost margin between final prices and costs (some of that margin may be due to monopoly rents which should not necessarily be included in access charges).

84. See example 2 of Box 2 of Chapter 1.

85. Article 3(3) of Regulation No. 2887/2000 of the European Parliament and of the Council of 18 December 2000 on unbundled access to the local loop. A few countries have set prices for ULL on a retail-minus approach in the interim. Spain writes: "When there is not enough information about costs and some kind of price

revision is considered necessary, access prices can be set temporarily on the basis of a retail-minus approach. This has been the case for indirect access to the local loop (through ADSL technology)".

86. Australia submission. This approach differs from the ACCC's TSLRIC model for PSTN services, where access prices differ according to these exchange service area classifications: CBD, Metropolitan, Provincial and Rural/remote. See Table A.2. The US also has geographically de-averaged local loop prices (by zones) within states, even though the retail service prices of an individual carrier are normally still averaged at the state level.

87. See OECD (2001b).

88. Oftel (2001), page 4.

89. Oftel (2001), page 5.

90. FCC (2001a), page 8.

91. See section 2.3.1. of this chapter.

92. The exception is calls to 0800 numbers.

93. Dutch submission.

94. ACCC (2000), page 17.

95. EC (2002), page 22-23.

96. A further complication is that retail charges for mobile operators are often highly non-linear – that is, they often include combinations of fixed charges, volume discounts and volume penalties. In the US, for example, mobile carriers have moved to offering large "buckets" of minutes of calls.

97. It is interesting to note that the EC Directives, which require cost-orientation and transparency for access charges for fixed network operators with SMP, do not require cost-orientation for mobile network operators with SMP. Article 7 of Directive 97/33/EC is limited in application to networks having SMP in the fixed public telephone network and the leased lines service.

98. It is proposed that in implementing the retail benchmarking approach the retail price movements will be determined using a yield methodology. Average retail prices will be calculated by dividing a mobile carrier's total mobile revenue from retail activities by the total number of its mobile minutes from mobile origination services in a given period. Effectively a revenue per minute figure (a proxy for average price) is derived that can be compared between periods in order to determine the retail price movements.

99. Australia submission.

100. In preparing this report a representative of Oftel argued that since mobile-to-mobile calls are routed by the fixed network they are, in practice, indistinguishable from fixed-to-mobile calls, so it is impossible to differentiate the termination charge on that basis.

101. "After passage of the 1996 Act, the FCC decided to have mobile-to-ILEC interconnection and reciprocal compensation determined pursuant to procedures found in sections 251 and 252. However, a CMRS industry group has petitioned the FCC to end the current system of reciprocal compensation and impose instead a 'bill and keep' regime. The FCC has requested comments on this request as part of the Intercarrier Compensation proceeding."

Appendix to Chapter 2

Table A.1. **Netherlands call origination and termination tariffs**

a) Tariffs for terminating access services for the period 1 July 2001 to 1 July 2002

	Average	Set-up	Conveyance		
			Peak	Off-peak	Weekend/night
National terminating	0.93	1.10	0.97	0.48	0.36
Regional terminating	0.72	0.78	0.76	0.38	0.28
Local terminating	0.54	0.59	0.57	0.29	0.21
National emergency connect	0.93	1.10	0.97	0.48	0.36
Regional emergency connect	0.72	0.78	0.76	0.38	0.28

Conveyance tariffs in € cents per minute (excl. VAT).
Set-up tariffs in € cents per call (excl. VAT).
Source: OECD.

Peak: Monday up to and including Friday 8:00 to 19:00;
Off-peak: Monday up to and including Friday 19:00 to 24:00;
Weekend/night: Night 00:00 to 08:00 and Saturday and Sunday 00:00 to 24:00.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

b) Tariffs for originating access services for the period 1 July 2001 to 1 July 2002

	Average	Set-up	Conveyance		
			Peak	Off-peak	WNT
Carrier selection national	1.22	1.43	1.27	0.63	0.47
Carrier selection regional	0.97	1.05	1.03	0.51	0.38
Carrier selection local	0.75	0.82	0.79	0.40	0.30
Carrier pre-selection national	1.24	1.45	1.29	0.64	0.48
Carrier pre-selection regional	0.98	1.07	1.04	0.52	0.39
Carrier pre-selection local	0.76	0.84	0.81	0.40	0.30
06760 Internet Connect regional (MIACO)	0.97	1.05	1.03	0.51	0.38
VPN connect 1-stage nat.	1.13	1.33	1.18	0.59	0.44
VPN connect 1-stage reg.	0.88	0.96	0.93	0.46	0.35
VPN connect 2-stage nat.	1.13	1.33	1.18	0.59	0.44
VPN connect 2-stage reg.	0.88	0.96	0.93	0.46	0.35
Paging connect national	1.13	1.33	1.18	0.59	0.44
Paging connect regional	0.88	0.96	0.93	0.46	0.35
National PN connect	1.13	1.33	1.18	0.59	0.44
Regional PN connect	0.88	0.96	0.93	0.46	0.35
800/90x connect national	1.34	1.57	1.39	0.69	0.52
800/90x connect regional	1.08	1.18	1.15	0.57	0.43

Conveyance tariffs in € cents per minute (excl. VAT).
Set-up tariffs in € cents per call (excl. VAT).
Source: OECD.

Peak:	Monday up to and including Friday 8:00 to 19:00;
Off-peak:	Monday up to and including Friday 19:00 to 24:00;
Weekend/night:	Night 00:00 to 08:00 and Saturday and Sunday 00:00 to 24:00.

06760 Internet Connect Flat Rate Service regional (FRIACO)	2.428

Tariff in euros per month per 2Mbit/s (excl. VAT).

90x Customer Billing Service	Tariff in € cents per call	1.43
	Percentage of the invoiced turnover	4.5%

Tariff excl. VAT.

Tariff per single number transfer (number portability)	33
Tariff per CPS change	1.94
Tariff for MDF access	11.93
TARIFF FOR LINE SHARING	6.30

Tariff in euros per month excl. VAT.

Table A.2. **PSTN originating and terminating access prices determined by the ACCC**

For The 2000-2001 Financial Year (Aust. cents)

Category	Flagfall	Per minute charge	Effective per minute charge
CBD	1.66	0.54	0.98
Metropolitan	1.74	0.96	1.43
Provincial	1.78	1.15	1.63
Rural/remote	1.99	2.22	2.76
National (weighted average)	1.76	1.05	1.52

Source: ACCC, A Report on the Assessment of Telstra's Undertaking for the Domestic PSTN Originating and Terminating Access Services, July 2000, p. 41.

Table A.3. **Maximum interconnection prices applicable in the Czech Republic in the year 2001**

According to CTO Decisions (CZK exclusive of VAT)

Category	Per minute charge	
	Peak	Off-Peak
One transit within a unified telephony circuit (UTO)	0.66	0.33
One transit to another UTO	1.08	0.54
Two transits	1.59	0.79

Source: Czech Submission.

Table A.4. **Rates for call origination and call termination in Germany**

€-cents/min., since 1 January 2002

	Peak rate (mon.-fri. 9-18 hours)	Off-peak rate (mon.-fri. 18-9 hours, sat., sun. and holidays)
Tariff gradient I (local)	0.65	0.44
Tariff gradient II (single transit)	1.07	0.71
Tariff gradient III (double transit)	1.86	1.22

Source: RegTP Press Release, 15 October 2001.

Table A5. **Interconnection rates in Mexico**

Type of service	$US cents per minute	Comments
Call origination and termination for long-distance service	1.32	This rate is 1.25 US cents plus an additional 5.7% for call attempts. This level has been maintained from the beginning of 2001.
Call termination on fixed networks for calls from other fixed networks	2.76	This tariff applies when traffic thresholds established in "bill and keep" systems are surpassed. This tariff is updated according to the CPI.
Call termination on fixed networks for calls from mobile networks	2.76	This tariff is updated according to the CPI.
Call termination on mobile networks for calls from fixed networks under Calling Party Pays	20.11	This tariff has remained the same since May 1999.
Call termination on mobile networks for calls from fixed networks under Receiving Party Pays	0.0	No interconnection tariff is paid in this case.

Source: Mexican Submission.

Table A.6. **Interconnection rates in Poland**

Category	Per minute rate (zloty/min)		
	T1 period – 8:00-18:00 weekdays	T2 period – 8:00-18:00 saturdays, sundays and holidays	T3 period – 18:00-8:00 all days of the week
Local interconnection termination/origination	0.032	0.024	0.016
Single-transit interconnection termination/origination	0.050	0.038	0.025
Double-transit interconnection termination/origination	0.068	0.051	0.034

Source: Poland submission.

Table A.7. **Prices of unbundled access to the local loop services and indicative prices of downstream retail services, exclusive of VAT (as of 1st Sept. 2001)**[1]

	Unbundled local loop access rental (connection fee)	Shared access rental (installation/connection fee, if applicable)	Retail rental services	Notes
Belgium	€ 11.33/month (PSTN/ISDN) € 13.96/month (broadband) (+/– 80 connection)	€ 4.54/month *(€ 86.51* connection)	PSTN € 13.5/month ISDN € 29.25/month ADSL > € 25.1/month	Prices in Belgium will have retroactive effect as of 31 December 2000.
Denmark	€ 8.23/month (first LL connection € 46.93; next orders to same MDF: € 16.36)	€ 4.12/month (€ *177* connection)	PSTN € 12.10/month ADSL > € 53.4/month	Investigation on cost orientation of prices not yet concluded.
Germany	€ 12.48/month (€ *92.6* connection)	–	PSTN € 10.94/month ISDN € 20/month[2] ADSL > € 28.5/month	ISDN is widespread in Germany[2]
Greece	€ 11.48/month (€ 123.38 connection)[2]	–	PSTN € 8.22/month ISDN € 13.21/month	the internal tie cable is included in the price[2]
Spain	€ 12.9/month (€ 103.92 connection)	€ 6.5/month (€ 103.92 connection)	PSTN € 9.27-11.4 /month (residential/business) ADSL > € 39.2/month	
France	€ 14.48/month (€ 107.9 connection)[2]	€ 6.10/month –	PSTN € 10.49/month ADSL > € 33.2/month	As determined by ART on 8 February 2001[2]
Ireland	€ 13.53/month (€ *119.73* connection)	€ 6.77/month € *178.93* connection	PSTN € 13.87/month	The Irish NRA has set these prices on 30 April 2001.
Italy	€ 11.62/month (PSTN/ISDN) € 14.15/month (broadband) (€ *89.86* connection)	–	PSTN € 10.69-14.62/month (residential/business) ADSL > € 69.7/month	
Luxembourg	€ 13.26 /month (PSTN/ISDN) € 15.79/month (broadband) –	–	PSTN € 16/month ISDN € 22/month	The Luxembourg NRA has not yet approved these tariffs, which were proposed by the incumbent operator.
Netherlands	€ 12.5/month (PSTN/ISDN) € 17/month (broadband) (€ *133.87* connection)	–	PSTN € 15.5/month ADSL > € 19.3 /month	This surcharge of 4.5 euro for broadband has not yet been approved by the NRA
Austria	€ 11.63/month *(until end 2001)* (€ *54.5* connection)	–	PSTN € 14.7 /month ADSL > € 33.8/month	
Portugal	€ 11.96/month (PSTN/ISDN) € 13.8/month (broadband) –	–	PSTN € 12/month ADSL > € 30/month	
Finland	< 3 km : € 12/month > 3 km : € 17.5/month (connection fee between € 5 and 320, average around € 200)	< 3 km : € 7.5 /month (no splitter), € 11/month (with splitter), > 3 km : € 10/month (no splitter), € 13 /month (with splitter) (connection fee: roughly between € 40 and 250).	PSTN € 11.76/month ADSL > 48.3/month (Sonera)	*
Sweden	€ 9.3/month (connection between € 156 and 168)	–	PSTN € 11/month ADSL > € 29.5/month	
UK	€ 16/month (connection fee € 142)	€ 9/month (connection fee € 208)	PSTN € 13.4/month ADSL > € 65/month	
Norway	€ 12.65/month for PSTN € 15.28/month for ISDN, € 21.24/month for broadband (connection fee between € 57 and 115)	€ 15.81/month (connection fee € 57.7)	PSTN € 15.3/month ISDN € 22.54/month	Member of the EEA

1. For Denmark, Sweden and UK conversion rates as of 2 April 2001 have been applied.
2. In Finland there are over 40 notified operators, so the data included in this table represent an average as indicated by the Finnish delegation at the ONP Committee of 7 February, and represents an industry average (prices are not determined by the NRA, the NRA checks whether the prices are cost oriented). The connection fees for full unbundled and shared access are estimated averages taken from the data provided by a majority of the notified operators in their responses to the questionnaire. The investigation on the cost orientation of the prices has not yet been concluded.

Source: OECD.

Table A.8. **Prices for local loops in Canada: retail prices and ULL prices**

	Band:	A	B	C	D	E	F	G
Bell Canada	Residential PES costs per month ($)	10.28	12.49	14.18	17.34	26.28	27.00	43.50
	ULL rates per month ($)	9.04	12.82	15.10	17.35	28.07	28.68	48.04
Island	Residential PES costs per month ($)	n.a.	17.67	17.93	n.a.	31.63	32.87	n.a.
	ULL rates per month ($)	n.a.	12.95	12.95	n.a.	30.02	31.66	n.a.
MTT	Residential PES costs per month ($)	15.97	n.a.	18.95	n.a.	28.40	28.35	n.a.
	ULL rates per month ($)	11.41	n.a.	14.95	n.a.	26.70	25.24	n.a.
MTS	Residential PES costs per month ($)	11.98	15.33	17.71	26.98	44/78	n.a.	85.80
	ULL rates per month ($)	6.04	13.59	17.53	24.70	44.41	n.a.	44.59
NBTEL	Residential PES costs per month ($)	n.a.	17.88	18.04	n.a.	30.00	24.07	n.a.
	ULL rates per month ($)	n.a.	12.52	14.69	n.a.	25.06	16.44	n.a.
NEWTEL	Residential PES costs per month ($)	n.a.	17.82	21.10	n.a.	31.53	30.98	37.02
	ULL rates per month ($)	n.a.	18.51	17.54	n.a.	26.61	27.50	40.04
TCI	Residential PES costs per month ($)	12.57	18.84	22.61	23.68	35.56	31.69	35.37
	ULL rates per month ($)	8.79	14.29	17.47	16.26	29.40	24.31	28.71
TCBC	Residential PES costs per month ($)	12.37	17.32	21.71	23.19	48.69	38.27	45.52
	ULL rates per month ($)	8.48	17.80	21.00	18.18	50.03	38.88	51.82
SASKTEL	Residential PES costs per month ($)	10.95	17.35	23.36	n.a.	45.70	38.53	53.87
	ULL rates per month ($)	10.09	16.62	23.27	n.a.	45.96	38.67	38.68

Source: Decision CRTC 2001-238-2, 7 August 2001 *www.crtc.gc.ca/archive/ENG/Decisions/2001/DT2001-238-2.htm*.

ISBN 92-64-10592-1
Pricing in Telecommunication

Chapter 3

Measurement of the Costs of Access Services

1. Introduction

Across the OECD, almost all telecommunications regulatory regimes require that the regulator, when setting access prices, has some consideration of the underlying "costs". For example, the 1996 US Telecommunications Act requires that prices for access to the services of incumbent operators be based on "costs".[1] Both the EU Directives and the WTO agreement on basic telecommunications services require that the prices of firms with significant market power be "cost-oriented".[2] This raises two fundamental questions: How do we measure the costs of a regulated firm? And what does it mean for prices to be cost-oriented or cost-based?

Of course, these questions are not unique to the problem of access pricing. They arise whenever a regulator controls prices in order to limit market power. If regulated prices are in some sense "too high" relative to costs, the regulator is not doing its job. If prices are in some sense "too low" relative to costs, the regulated firm will fail to make necessary investment. For better or for worse, therefore, regulatory authorities are led to ask the question: "how much does it cost to provide the regulated services in the given period"?

2. The cost allocation problem

Attempts to answer this question come up against a fundamental problem. If the provision of access services requires an investment in a sunk asset, a meaningful measure of the "cost" of providing the regulated services in the given period requires an allocation (or "amortization") of some part of the original cost of purchasing the asset to that period. For example, suppose the provision of a given service requires an investment of $1 million in an asset that lasts exactly ten years and suppose the regulator wishes to determine the "cost" of providing services in year 5. What share of the $1 million investment should be attributed to year 5?

This problem can be viewed as just another example of the problems that arise when there are costs which are common to two or more services. When there are costs which are shared between two services X and Y, the question "how much does it cost to produce service X?" requires an allocation of the common costs to service X. The cost of a sunk investment in an asset can be viewed as a form of common cost. This cost is common to all the services produced with the asset over the life of the asset. The problem of determining

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

the cost of providing services in any given period requires an allocation of the cost of the asset to that period.

Regulators have for many years attempted to come up with "estimates" of the cost of individual services by allocating the common costs according to various methodologies (*e.g.*, in proportion to the quantities of different services sold). These approaches are collectively known as "fully allocated cost" or "fully distributed cost" approaches. They have been criticised by economists as arbitrary and yielding economically meaningless results. In the same way, most approaches for allocating (or amortizing) the costs of fixed investments over the life of that investment also follow arbitrary rules of thumb. Most of the time, these are the inter-temporal equivalent of fully-allocated cost.

How should the costs of sunk investments be allocated? From the perspective of overall efficiency and welfare, all that matters is the set of prices ultimately charged to users and consumers. Cost allocation only matters insofar as it affects prices. The question of how should we allocate costs then becomes a question about the optimal structure of prices. As we will see later in this chapter, the structure of the prices of a regulated firm might be chosen in different ways – for example, they might be chosen in a way which maximises overall welfare, or in a way which minimizes the scope for inefficient entry.[3]

It is worth noting that this basic problem of cost allocation has important implications for the detection of monopoly profits. Measurement of monopoly profits requires information on both revenues and costs. As we have emphasized, the measurement of costs in any one period requires an allocation of costs. In most cases the regulator has discretion over how it chooses to allocate these costs. The measurement of the "profit" a firm is making in any one period is therefore largely arbitrary. Estimates of monopoly rents based on a comparison of current revenue and some measure of current costs should therefore be treated with caution.[4]

3. Review of basic principles, depreciation and the regulatory asset base

In order to better understand the problem of cost allocation, let's return to first principles. We will focus on the amount a regulated firm earns in profit each period after paying for all the costs that are directly attributable to that period. We will call this amount the firm's "earnings" (in the sense of the accounting term "Earnings Before Interest Taxes, Depreciation and Amortisation" or "EBITDA"). These earnings, or "operating profits", are the amount that the firm is able to return to its investors as compensation for their investment.

When a firm must make a sunk investment, what does it mean for a firm to be earning sufficient revenue to cover its costs? The answer from economic theory is clear: A firm is earning sufficient revenue to just cover its total costs (or, equivalently, a firm is earning no monopoly profits) if its future earnings over the life of the firm are sufficient to cover the costs of its investment. Or, technically, a firm is earning sufficient revenue to cover its costs if and only if *the net present value of its stream of earnings less its capital investment over the life of the firm is zero.*

In other words, if the firm is to receive sufficient revenue to justify its investments (but no more) the path of revenue over the life of the firm must be such that the present value of the stream of net payments to investors is just zero. We can now give a concrete definition to the term "cost based" – a set of prices is cost based if it leads to a path of revenue such that, over the life of the firm, the present value of the stream of net payments to investors is just zero.[5]

We can express this mathematically as follows. Suppose that the firm has a life of N periods and suppose that the path of earnings over the life of the firm is $E_1, E_2, \ldots, E_N$. Similarly, let I_0 be the initial capital requirements of the firm and $I_1, I_2, \ldots, I_N$ be the on-going capital expenditure of the firm (on new assets, renovation or extension of existing assets, rebuilding, or maintenance). Finally, let the opportunity cost of the firm's capital be given by r (for simplicity, throughout this chapter we will assume that this cost of capital is constant). The given path of earnings allows the firm to just cover its capital costs if and only if:

$$\sum_{s=1}^{N} \frac{E_s}{(1+r)^s} = \sum_{s=1}^{N} \frac{I_s}{(1+r)^s} + I_0 \qquad \ldots(1)$$

A regulator which seeks to impose cost-based prices seeks to control the earnings of the firm in such a way as to ensure that, at least on average, over the life of regulated firm, the present value of the earnings of the firm is equal to the present value of the capital costs. Or, expressed in mathematical notation, the regulators job is to choose a path of regulated prices which satisfy equation 1, at least on average. If the regulated firm expects to earn more in the future (in present discounted value terms) than it expects to spend on capital investment, the regulated firm is earning excess returns or monopoly rents. On the other hand, if the regulated firm expects to earn less in the future (in present discounted value terms) than it expects to spend on capital investment, it will fail to make the investment.

3.1. The asset base, depreciation, etc.

How do regulators go about choosing a path of prices which satisfy equation 1? In practice, almost all regulatory authorities make use of a device

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

that transforms the problem into one which is conceptually a little simpler, although mathematically identical.

This transformation of the basic problem involves the introduction of a new variable, which we will call the "asset base". The asset base is also sometimes referred to as the "rate base" of the firm's assets.[6]

The "asset base" can be viewed as an amount that the owners of the regulated firm have "in the bank". Each period this amount increases by the interest paid by the bank and by any additional "deposits" in the form of any new capital investment (whether in the form of creation or purchase of new assets, renovation, rebuilding or refurbishment of existing assets, or maintenance) and decreases by any "withdrawals" in the form of a payout (in the form of earnings) to investors. The asset base can therefore be seen as the amount "owed" to the regulated firm by the regulator. At any point in time if the regulated firm were shutdown and its owners given a payment equal to the asset base (*i.e.*, if there was a "withdrawal" of all the remaining funds in the "bank"), they would be fully compensated for the costs of investment they had incurred up to that point in time.

It is straightforward to check that the basic task of the regulator (as summarised above in equation 1) is equivalent to ensuring that the asset base declines to zero at the end of the life of the regulated firm. If the asset base becomes negative at the end of the life of the regulated firm, the regulated firm has earned more than enough to cover its costs (*i.e.*, has earned some monopoly rents). If the asset base remains positive at the end of the life of the regulated firm, the regulated firm has not recovered all its costs.

This can be expressed mathematically. The problem of the regulator can be expressed as the choice of a path of the asset base over the life of the firm $K_0, K_1, \ldots, K_N$ such that *a)* in the first period the asset base is equal to the size of the initial investment (*i.e.*, $K_0 = I_0$); and *b)* the asset base is zero at the end of the life of the regulated firm (*i.e.*, $K_N = 0$); and c) in each period the asset base evolves as follows:

$$K_t = (1 + r)K_{t-1} - E_t + I_t \qquad \ldots(2)$$

Having introduced the notion of the asset base, the basic task of the regulator can be re-expressed in the following way: The regulator's task is to choose a path of the asset base over the life of the firm in such a way that the initial asset base is equal to the initial investment, the asset base is zero when the firm no longer exists and, in each period the earnings allowed to the regulated firm satisfy equation 2. Any path of the asset base which satisfies these rules, ensures that the net present value of payments to investors over the life of the firm is zero.

Are there any constraints at all on the path of the asset base that the regulator can choose? In practice, the path of the asset base is limited by the upper and lower bounds on the level of earnings that the firm can earn in each period. If we assume that the regulated firm cannot sell for a price lower than zero (*i.e.*, the firm cannot charge less than simply giving away the service), the earnings cannot be lower than the loss of the operating cost. At the same time the earnings in any one period cannot usually be larger than the monopoly profits for the service in that period. These upper and lower limits on earnings translate (through equation 2) into upper and lower limits on the change in the asset base from one period to the next.

If there was a liquid and developed second-hand market for the assets of the regulated firm which is independent of the actions of the regulator, the path of the asset base could be taken from the market price for identical assets selling in the second-hand market.[7] The path of market prices of the asset would in turn (via equation 2), determine a path of earnings of the firm. In fact, if there is a liquid and developed second-hand market for the assets of the regulated firm, the regulator's task becomes very simple. Since there is no longer any discretion involved in determining the appropriate level of earnings, such assets need not even be included within the asset base of the regulated firm. Instead, such assets can be handled as though they are leased. Leased assets are not included in the asset base, but are treated as operating costs.[8]

Unfortunately, the existence of a liquid second-hand market independent of the actions of the regulator, is the exception rather than the rule. In most cases, there is simply no meaningful second-hand market and, if there were, the price of the assets in that market would be fundamentally affected by the decisions of the regulator as to the level of the regulated prices. Any attempt to set regulated prices on the basis of the value of second-hand assets in such a market raises the problem of circularity. This approach is of little use in allocating the capital costs of most assets of most regulated firms.

Whatever path of the asset base is chosen, if the regulator is doing its job perfectly, the level of the asset base is at all times equal to the net present value of all future payments to investors. In other words, the asset base is, at all times, equal to the value at which the regulated firm would sell in an efficient market.

This observation has some important potential implications. If the regulator observed that the price of the regulated firm's shares would value the firm at a price significantly above its asset base, this implies that either *a)* the regulator is using a cost of capital that is too high; *b)* the regulator is allowing earnings which are too high; *c)* the regulated firm has costs that are lower than those recognised by the regulator; or *d)* the regulated firm has earnings which are not being counted by the regulator. Conversely if the regulator observes that the price of the regulated firm's shares value the firm at a price below its asset base, this implies that the market believes that either *a)* the regulator is using a cost of capital that is too low; *b)* the regulator is allowing earnings which are too low; *c)* the

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

regulated firm has operating costs or capital costs that are higher than recognised by the regulator; or *d)* the regulated firm has earnings lower than the regulator believes. This suggests that the market price of the regulated firm's shares might be used as a kind of indicator of the performance of the regulator. Ehrhardt (1994) writes:

> "In fact, some researchers argue that regulators should set prices so that the market and book values of equity are equal, that this is equivalent to setting the allowable rate of return equal to the market-determined weighted-average cost of capital. Whether regulators actually pursue this goal is debatable, but it is true that book and market values of equity for utilities are often very similar."[9]

3.2. Depreciation

Although (as we will see shortly) some regulatory authorities do focus directly on the path of the asset base, the single most common approach of regulators involves defining one additional term: "depreciation".

It is important to note that the word depreciation has two possible meanings. These meanings are often confused, obfuscating rather than clarifying the essential ideas. The two possible meanings of depreciation are: *a)* the *physical* notion of the decline in productive capacity of an asset (either due to a reduction in its capacity or an increase in its operating costs) and *b)* the *accounting* or economic notion of allocation of the cost of an asset to a particular period (or, equivalently, the decline in the value of an asset).

Economists have occasionally proposed that these two concepts be distinguished using two separate labels. For example, some have argued that the word "depreciation" be reserved for the loss of productive efficiency of an asset, while the word "amortisation" is reserved for the accounting concept of spreading the cost of an asset over several periods. In this chapter, the word depreciation will refer to the accounting notion.[10] Box 11 discusses in more detail the different meanings of the word depreciation.

Suppose we define the level of depreciation in a given period to be the change in the value of the asset base over that period plus any new capital expenditure in the period. In this case the problem of the regulator can be expressed as finding the path of the depreciation over the life of the regulated firm subject to the constraints that *a)* the total amount of depreciation must add to the total capital expenditure and *b)* in each period the earnings of the regulated firm must be equal to the cost of capital times the asset base plus the depreciation.

Using mathematical notation, the basic task of the regulator can be re-expressed as follows: The regulator's task is to choose a path of depreciation $D_1, D_2, \ldots, D_N$ such that *a)* the path of the asset base evolves according to

$K_t = K_{t-1,}\ I_t - D_t$; *b)* the path of earnings is given by $E_t = rK_{t-1,}\ D_t$; and *c)* the sum of the depreciation over the life of the regulated firm satisfies:

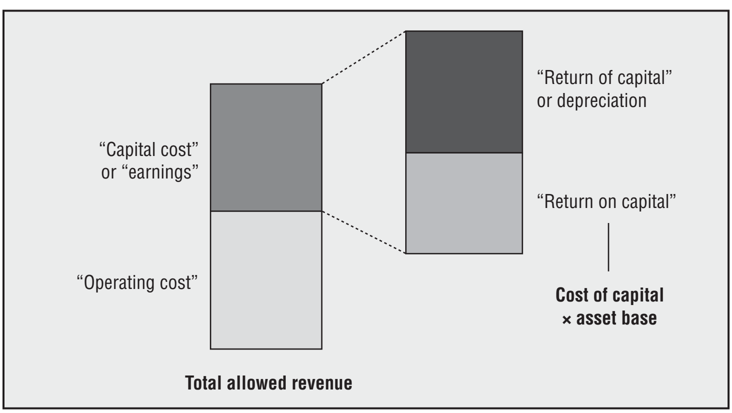
$$\sum_{s=1}^{N} D_s = \sum_{s=1}^{N} I_s + I_0 \qquad \ldots(3)$$

When the regulator's problem is expressed in this way, the earnings allowed to the firm each period is equal to the sum of two terms: The first term, called the "return on capital" is equal to the cost of capital times the size of the regulatory asset base at the end of the previous period. The second term is called the "return of capital" and corresponds to what we are calling here the "depreciation".

This way of expressing the task of the regulator, which involves calculation of *i)* the regulatory asset base, *ii)* depreciation (related to the change in the value of the asset base) and *iii)* earnings (given by the return on capital and return of capital) underlies the approach used by virtually all regulatory authorities. It is sometimes known as the "building block" approach.[11]

Figure 22 summarises diagrammatically the building block approach to determining the total cost (or revenue) allowed to the firm in any given period. The total cost is the sum of the operating costs and the capital costs. The capital costs are given by the return on capital (equal to the cost of capital times the asset base) and the return of capital (given by the depreciation).

Figure 22. **The standard "building block" approach**



Source: OECD.

3.3. Variations

There exist variations of the basic "building block" approach that are nevertheless consistent with the basic requirement that the regulated firm earn sufficient revenue to cover its costs (as in equation 1). These variations are possible because the concepts of "asset base" and "depreciation" are

Box 11. **Depreciation *vs.* amortisation**

As mentioned in the text, discussions of depreciation are made significantly more complicated by confusing two distinct notions – the technical or physical notion of the loss in productive efficiency or productive ability of an asset and the accounting or economic notion of the change in the "value" of an asset or (equivalently) the allocation of the cost of the asset over its useful life.

This confusion is apparent, for example, in the definition of depreciation offered by the American Gas Association. In its glossary of terms, the American Gas Association defines depreciation as:

"Return of investment through inclusion in cost of service (and rates) of a pro rata part of the cost of property, calculated to spread the total investment cost over a certain period of time or number of units that measure the useful life of the investment."

This portion of the definition is a good reflection of the accounting notion of depreciation, but the American Gas Association definition goes on to state that:

Depreciation (in the Code of Federal Regulations) is to reimburse the company for "… the loss in service value not restored by current maintenance, incurred in connection with the consumption or prospective retirement of gas plant in the course of service from causes which are known to be in current operation and against which the utility is not protected by insurance. Among the causes to be given consideration are wear and tear, decay, action of the elements, inadequacy, obsolescence, changes in the art, changes in demand and requirements of public authorities, and, in the case of a natural gas company, the exhaustion of natural resources".

Concepts such as "wear and tear" or "action of the elements" relate more closely to the physical concept of depreciation – the loss in productive efficiency of the asset, either due to declining capacity or increased maintenance costs. The complete definition confuses two distinct ideas.

Some economists have suggested using different labels for these two concepts. For example, the word depreciation could be reserved for the physical deterioration of an asset, while another word, such as "amortisation" could be used to refer to the accounting concept of depreciation. The American Gas Association defines amortisation as follows:

> The gradual extinguishment (or accumulated provision or reserve therefore) of an amount in an account by pro rating such amount over a predetermined period.

From the perspective of the model set out in this chapter, the two notions of depreciation can be distinguished in the following way. The physical notion of

Box 11. **Depreciation *vs.* amortisation** *(cont.)*

depreciation affects only the link between a given price and a given level of earnings (*i.e.*, operating profits) – and possibly the maximum or monopoly earnings in a period. In contrast, the accounting notion of depreciation relates only to the choice of the allowed level of earnings. Since, in this chapter we will be focused primarily on how the allowed level of earnings should be chosen, it is the second, accounting notion of depreciation, on which we will focus.

merely useful mathematical devices that allow the basic task of the regulator to be expressed in a simpler way.

For example, the regulator might consider that the size of the regulatory asset base should be inflated each period by the rate of inflation (this might be linked to a view that, due to inflation, the replacement cost of the asset is increasing over time). Suppose the asset base is scaled up by inflation, so that the path of the asset base evolves according to $K_t = (1 + i)K_{t-1} - I_t + D_t$ where i is the inflation rate. This is still consistent with the basic constraint on the path of the asset base given by equation 2 provided that the "return on capital" is calculated using a *real* interest rate (*i.e.*, $r - i$) rather than the nominal interest rate (r) – *i.e.*, if the path of earnings is given by $E_t = (r - i)K_{t-1} + D_t$.[12] This form of "adjustment for inflation" has no necessary effect on the path of earnings of the regulated firm.

As another example, the regulator might use a definition of depreciation that reflects the *total* change in the value of asset base (rather than the change in the value of the asset base before new capital expenditure as above). The path of the asset base under this new approach evolves according to $K_t = K_{t-1} + D_t$. This is entirely consistent with the basic constraint given by equation 2 provided that the path of earnings is given by the sum of three terms: the return on capital (using the nominal interest rate), the return of capital (given by the depreciation) and a third term corresponding to capital expenditure in the period, *i.e.*, $E_t = rK_{t-1} + D_t + I_t$.

Since what we are calling earnings here is equal to revenue less operating costs, we can re-arrange this last expression to show that in each period the maximum revenue allowed to the regulated firm can be expressed as the sum of four terms: *a)* operating expenditure; *b)* capital expenditure; *c)* the rental cost of capital employed in the period (*i.e.*, "return on capital") and *d)* a term corresponding to the change in the value of the asset base. Note that in this case the sum of the total depreciation over the life of the firm is just the initial investment (compare this with equation 3):

$$\sum_{s=1}^{N} D_s = I_0 \qquad \ldots(4)$$

 PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

3.4. The task of the regulator (again)

Remember that we are assuming that the basic task of the regulator is to set cost-based prices – that is, to choose a path of earnings over the life of the firm in such a way that the present value of the net-payments to investors is zero. In this section we have seen that this basic task can be re-expressed in two different ways:[13]

1. First, the regulator could choose a path of the asset base $K_0, K_1, \ldots, K_N$ such that in the first period the asset base is equal to the size of the initial investment (*i.e.*, $K_0 = I_0$); and the asset base is zero at the end of the life of the regulated firm (*i.e.*, $K_N = 0$); and in each period the asset base evolves as follows:

   $K_t = (1 + r)K_{t-1} - E_t + I_t$

2. Second, the regulator could choose a path of depreciation $D_1, D_{2,} \ldots, D_N$ such that the path of the asset base evolves according to $K_t = K_{t-1,}\ I_t - D_t$; the path of earnings is given by $E_t = rK_{t-1} + D_t$; and the sum of the depreciation over the life of the regulated firm satisfies:

$$\sum_{s=1}^{N} D_s = \sum_{s=1}^{N} I_s + I_0$$

These three approaches are equivalent. Given the path of any one of: earnings, the asset base or depreciation, it is straightforward to find the corresponding path of the other two. At the same time, as long as the regulator is committed to its task of allowing the regulated firm sufficient earnings to cover its costs and no more, the regulator can *only* choose one of these three paths. It cannot, for example, simultaneously choose the path of the asset base *and* the path of depreciation. Once the path of the asset base is chosen it determines the corresponding path of depreciation. Conversely when the path of depreciation is chosen it determines the corresponding path of the asset base.[14]

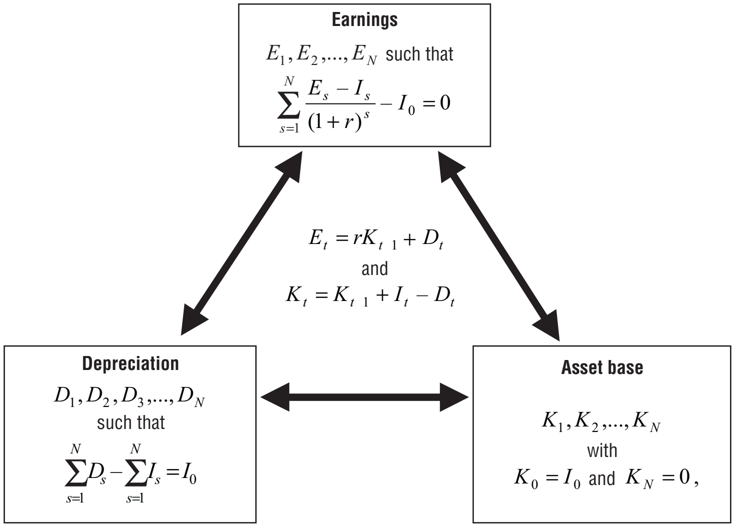
This is summarised in the following diagram (Figure 23).

4. The historic cost and depreciated replacement cost approaches

4.1. The original cost or historic cost approach

As we have seen many times now, the basic task of the regulator is to choose a path of earnings over the life of the firm that has a present value of the net payments to investors equal to zero. The single most common way of addressing this basic task is through a choice of the path of depreciation. As set out in the previous section, the regulator must choose an amount of depreciation each period in such a way that the total depreciation over the life of the firm is equal to the total capital expenditure of the firm.

Figure 23. **Equivalent ways of expressing the basic task of the regulator**



Source: OECD.

The simplest way to do this is through so-called *uniform* or *straight-line* depreciation. Under this approach, the depreciation each period is equal to the initial capital expenditure divided by its life. For example, if an asset costs \$1 million and lasts 20 years, the straight-line path of depreciation will be \$50 000 per year. If in the 10th year of the life of this firm, an additional capital expenditure of \$50 000 is required, with a "life" of 5 years, the existing path of depreciation is increased to \$60 000 in the 10th to 15th years. As an illustration, Figure A.1 in the appendix illustrates the earnings path that results from the straight-line depreciation of an asset that lasts ten years and is replaced at the end of its life. As can be seen in Figure A.1, straight-line depreciation leads to a declining path of earnings (at least in periods between major capital expenditure).

Another way of choosing the path of depreciation is known as *sum of the years' digits*. Compared to uniform depreciation, this approach involves higher levels of depreciation in the early years of an asset's life, and lower levels of depreciation later. Figure A.2 in the appendix illustrates the earnings path resulting from sum-of-the-years'-digits depreciation.[15]

In years when there is no capital expenditure, both uniform and sum-of-the-years'-digits depreciation lead to a declining corresponding path of earnings. One consequence is that the corresponding earnings path will be discontinuous when the asset reaches the end of its useful life and must be replaced unless the cost of replacing the asset is also declining. Typically, the path of earnings will have large "jumps" when new capital expenditure is required, especially when the cost of new capital expenditure is rising. This can be seen in both Figure A.1 and Figure A.2.

These "jumps" in allowed earnings will typically imply a discontinuity or "jump" in the regulated prices. This can result in what is known as "tariff shock" or "rate shock", as customers and consumers react angrily to a sudden large increase in price.

The size of the "rate shock" will clearly depend on a number of factors, such as the magnitude of the capital expenditure required both in absolute terms and relative to the initial capital expenditure. Where the regulated firm is comprised of a number of small assets, the replacement of any one of these assets would not typically involve significant rate shock. On the other hand, where the assets of the regulated firm are "lumpy" and where the replacement costs have increased significantly over time, the resulting rate shock might be very large. This problem of rate shock seems to be a real concern of regulators.

Without abandoning the original cost approach entirely, rate shock might be reduced by adjusting the path of depreciation given by straight-line depreciation or sum-of-the-years'-digits. For example, a regulator might allow higher levels of depreciation in the years leading up to a major new capital expenditure, in order to lift required earnings (and therefore prices) in anticipation of an earnings (and prices) rise to come. Such an approach might involve a degree of "pre-financing" of the new investment.

4.2. The depreciated replacement cost approach

The rate shock problem can be viewed as arising from the disparity between the historic cost of the investment and the current or replacement cost of the firm's assets. In a time of general inflation, the replacement cost of the regulated firm's assets may be much higher than the historic cost. Consequently, when the life of the existing assets expires, earnings must increase significantly to finance the new investment. Some regulators therefore choose to focus not on the path of depreciation, but on the path of the asset base, to ensure that, at all times, the firm's asset base reflects (in some way) its current replacement cost.

But which "replacement cost" should be used? Is it the cost of replacing the existing asset with an identical (but brand new) asset? Or the cost of obtaining the same set of services using a modern replacement asset (which may have a longer life or lower costs)? Or the cost of replacing the asset with an asset which can provide a higher set of services if there is demand for those services? How many of the assets of the incumbent will be replaced at once? The whole lot? – which would allow for a complete re-configuration and re-optimisation of the network, or just a subset? – in which case the replacement would only allow for a partial re-configuration or re-optimisation, at best? These issues are largely beyond the scope of this paper. In the next sections we will focus on the case where the relevant replacement cost is the cost of building or buying a brand new "modern equivalent asset". (The extent to which this asset is "optimised" will be discussed below). This is sometimes called the "Optimised Replacement Cost" or "ORC".

One simple way to reflect changes in the replacement cost in the regulatory asset base is to scale up the value of the asset base by the rate of inflation each period. This approach is also called "trended original cost" or "real straight line" depreciation. In this context the asset base evolves as follows: $K_t = (1 + i)K_{t-1} - D_t$. The resulting path of earnings is illustrated in Figure A.3.[16]

The trended original cost approach uses a very simple adjustment to the path of the asset base in an attempt to bring it more into line with the replacement cost. A slightly more sophisticated approach is to directly link the level of the asset base with the current replacement cost.

Under this approach the asset base is set equal to the current price of a new equivalent asset "depreciated" in some way to reflect the fact that the existing assets have a shorter remaining useful life than a new asset.

It is common to "depreciate" the replacement cost using "straight-line depreciation". Under this approach, in each period the asset base is set equal to the current replacement cost, scaled down by a factor reflecting the number of years of remaining life of the existing asset. For example, if a new asset costs $100 000 and the existing assets have a useful remaining life of, say, 6 years out of an original life of 10 years, the asset base is set equal to $60 000.[17]

We will refer to this approach as the "depreciated replacement cost" (or "depreciated optimised replacement cost" or "DORC") approach. A possible path of earnings when the asset base is calculated using depreciated replacement cost is illustrated in Figure A.4.

Does this approach solve the problem of "rate shock"? As we can see from Figure A.5, the answer is clearly No. In fact, as long as the replacement cost is increasing at a rate faster than the interest rate, the path of earnings given by depreciated replacement cost will be increasing at a rate *faster* than the replacement cost. Conversely, (as in Figure A.4) if the replacement cost is increasing at a rate less than the interest rate, the path of earnings will be increasing at a rate *slower* than the replacement cost. Since the interest rate can never be negative, if the replacement cost is declining over time the path of earnings will always be declining faster than the replacement cost. In all of these cases, there is inevitably a jump in earnings when the underlying asset is replaced. Under the depreciated replacement cost approach the path of earnings is continuous only when the interest rate equals the rate of increase in replacement cost.

Does the depreciated replacement cost approach do a better job at reducing rate shock than simple straight-line depreciation? As Figure A.5 shows, the depreciated replacement cost approach may do a somewhat better job at ensuring a continuous path of earnings than simple straight-line depreciation when the replacement cost is increasing. In this case, the path of earnings under DORC is at least increasing (compared to straight-line depreciation where it is always declining). On the other hand, when the replacement cost is declining the relative advantages of depreciated replacement cost are less clear.

In Figure A.4 and Figure A.5 the replacement cost followed a smooth continuous path. Discontinuity in the path of earnings is also likely when the path of replacement cost is not smooth, but is accelerating or changing direction.

5. Historic cost *vs.* replacement cost, optimisation and cost models

In the previous sections we looked at two different approaches for determining the path of depreciation of a regulated firm. The first approach was to choose the depreciation directly (on the basis of the historic cost). The second approach was to, instead, choose the path of the asset base, with depreciation then determined implicitly as the difference in the asset base between two periods.

These two approaches have several important similarities:

- First, because both approaches maintain the "building block" process over time, both ensure that the regulated firm will be neither under-compensated nor over-compensated over its life.
- Second, under both approaches the effect of the particular cost allocation chosen on prices is downplayed or ignored entirely. As a consequence, the resulting prices may be undesirable from a welfare point-of-view.
- Third, both approaches are, in some sense, equivalent. Any path of earnings that results from a choice of a path of depreciation may also arise from a choice of the path of the asset base and *vice versa* .
- Lastly, as we have seen, neither approach can guarantee that the path of earnings will be continuous. As a rough guide, when the replacement cost is increasing the depreciated replacement cost approach is likely to lead to a smoother path of earnings, especially when the rate of increase of replacement cost and the cost of capital are not too different.

In addition, the two approaches do not differ in their use of "optimisation". It is usually accepted that in order to maintain incentives on the regulated firm to control its costs it is necessary to impose some kind of controls on the level of capital expenditure by the regulated firm allowed by the regulator. These controls may take the form of explicit review and audit of the capital expenditure decisions of the regulated firm, or by basing the compensation of the firm on the basis of the capital expenditure decisions given by an objective benchmark – either a cost model, or the expenditure decisions of related "benchmark" firms. However, these controls can be (and often are) applied in both the historic cost and the depreciated replacement cost approaches.

However, the depreciated replacement cost approach does introduce some new problems which are addressed in this section. These problems raise questions as to the extent to which the "pure" depreciated replacement cost approach can be used in practice.

5.1. *Refurbishment expenses*

Recall that under a (pure) depreciated replacement cost approach the asset base of the regulated firm is set equal to the depreciated cost of purchasing or constructing a brand new modern equivalent asset. In the typical case, some mechanical approach to depreciation (such as straight-line depreciation) is used to reduce the cost of a brand new asset to a value reflecting the short remaining life of the existing asset.

In this chapter until now the only capital expenditure that has been required is the complete replacement of the underlying asset. In fact capital expenditure can be divided into two categories: *a)* capital expenditure which creates, enhances or extends the regulated firm's asset, on the one hand, and *b)* capital expenditure which merely maintains, renews or refurbishes an existing asset on the other. In the case of capital expenditure which creates, enhances or extends the asset, since the expenditure is changing the scale or facilities of the underlying asset the expenditure will have some impact on the corresponding replacement cost of an equivalent asset. Investment in a new gas pipeline or an extension to a cable television network, for example, will have the effect of increasing the cost of replacing the entire network. On the other hand expenditure which merely renews or refurbishes an existing asset, although a legitimate capital expenditure, does not necessarily affect the cost of buying an modern equivalent asset – replacing the engine in a used car may be a necessary expenditure but it does not change the cost of buying a new car.

A problem arises for the second category of capital expenditure when the asset base is based on the depreciated replacement cost. Since, under the DORC approach, the asset base is determined entirely by the replacement cost, such expenditure is not incorporated into the asset base. Instead, such capital expenditures are treated as an operating expense and are therefore depreciated 100% in the period in which they are incurred. This means that refurbishment expenses have to be borne entirely by earnings in the period in which they are

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

incurred. Obviously, for large refurbishment expenses, this could lead to large jumps earnings and prices.

There are at least two possible solutions to this problem:

1. One approach is to view the purchase of an asset not as a single one-off cost, but as a commitment to incur a stream of costs over time. The cost of operating a passenger plane is not simply the cost of purchasing the plane but the cost of regular overhauls and servicing. The cost of owning a house includes the cost of regular repairs, repainting and maintenance. All of these on-going costs are annualised into the regular costs of leasing a plane or a house.

   Under this view, the asset base should not be based on the cost of purchasing or constructing a modern equivalent asset, but rather the present discounted value of the forecast future stream of costs associated with a modern equivalent asset purchased or constructed today. Subsequently, refurbishment expenditures (to the extent that they were correctly forecast at the outset) need not be taken into account at the time they are incurred (they have already been capitalised into the asset base).

2. The second possible approach is to divide the asset of the regulated firm up into its constituent parts which each have their own replacement cost, useful life and rate of depreciation. The total DORC valuation is then the sum of the DORC valuations for the underlying assets separately. A major refurbishment expenditure is then viewed as the replacement of one asset comprising the overall asset of the regulated firm (such as the replacement of the "paint").

Note that in both of these approaches the DORC valuation is no longer related in a simple way to the cost of building or purchasing a modern equivalent asset.

5.2. Problems arising from economics of scale and scope in capital expenditure programmes

Another fundamental problem with the use of the DORC methodology arises when there are economies of scale or scope in the construction of new capital assets. Some form of economies of scale and scope are quite common. For example, if a trench must be dug to lay a telecommunications cable, it often makes sense to lay additional cable in anticipation of possible future demand than to run the risk of having to dig up the trench again in the future. Alternatively, if a network must be built that can provide service A today, but it is anticipated that service B will be needed in the future, it often makes sense to install a network capable of providing both A and B at the outset rather than run the risk of having to upgrade the network later.

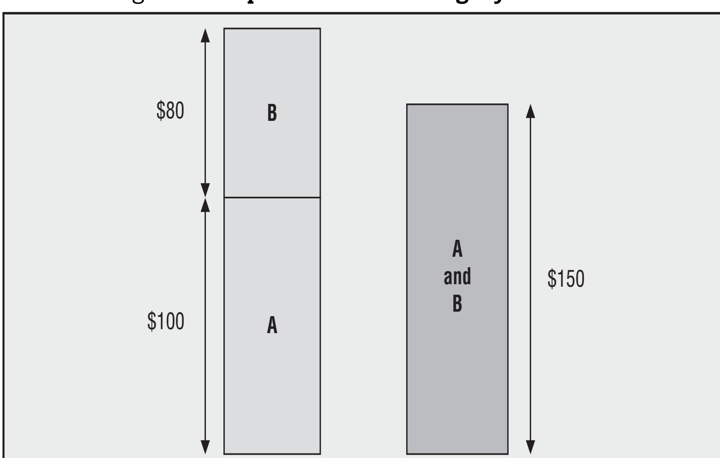
When there are economies of scale or scope in capital expenditure programmes the history of the industry matters. In particular, if service A was needed at some time in the past and service A+B is needed today, it may be that the most efficient way of meeting that demand is to build a network capable of providing service A at the outset and then to upgrade it to provide services A and B, even though it would be cheaper today to build a network from scratch which provided both services.

This is relevant for the DORC methodology because the (pure) DORC methodology only allows the regulated firm an asset base based on the least cost of providing the services required today. If the regulated firm actually incurs higher cost in providing those services (because the services were added incrementally over time as demand materialised) those additional costs must be expensed 100% in the year in which they are incurred.

5.3. Problems with legacy networks

This can be illustrated with an example. Suppose that, at time zero, a network is created to provide services A. At that time the probability of service B being required is sufficiently low that it is not considered worthwhile to build a network capable of providing both services A and B. This network lasts two regulatory periods. The efficient cost of the network is $100. Suppose that after one regulatory period, demand for B materialises. The network can be upgraded to also provide service B at a cost of $80. The stand-alone cost of providing service B is $100, so it is cheaper to upgrade the existing network than to provide the new service with a new facility. However, the efficient cost of new "greenfields" facility to provide both services is only $150. For simplicity assume that the cost of capital is zero.

Figure 24. **Optimisation and legacy networks**



Should the regulated firm be allowed an asset base of $180 after the upgrade (the original $100 plus the upgrade cost of $80) or just $150 (the efficient cost of providing A and B)?

Source: OECD.

Suppose that the regulator adopts a DORC methodology – that is, it sets the regulatory asset base of the regulated firm on the basis of the current efficient replacement cost of the asset. This would allow an asset base of only $150 in this example. If the firm is to be induced to take the additional $80 of investment, the additional $30 must be expensed 100% in the current period.

The additional $30 of expenses will contribute to instability in prices which, as a form of rate shock, is usually considered undesirable. The only alternative is to fully capitalise the additional expenditure – that is, to allow the regulated firm an asset base of $180. In this case however, since the asset base is not set on the basis of the cost of building or purchasing a modern equivalent asset, the regulated firm is no longer using a (pure) DORC methodology.

5.4. The degree of optimisation in telecommunications cost models

This issue has arisen particularly in the telecommunications industry in fixing the framework for telecommunications cost models. In the telecommunications industry, in particular, it has become very common to make use of computerised cost models as the basis for determining an objective standard for cost reimbursement of a regulated firm.[18]

Regulators typically asset that these cost models are designed to estimate the least cost of a "modern equivalent asset" – that is, the efficient cost of a network providing services equivalent to the existing network – given the geographical locations of customers, a given level of demand for telecommunications services and a given set of prices for the key inputs used to provide telecommunications services (switching, cabling, trenches and so on). As such, these cost models are one form of the DORC methodology.

One key parameter into these cost models is the degree of optimisation – the extent to which the model considers alternatives ways to configure the network in order to provide the required services. In particular, the models may need to determine, for example, the optimal location of switching centres and the optimal mix of switching and transportation in the network. One key question that has emerged is whether to optimise over the location of both *a)* the links between customers and switches and the locations of switches or just *b)* over the links between customers and switches (taking the locations of switches as fixed). The former approach is known as the "scorched earth" or "greenfields" approach as it allows all the features of the hypothetical replacement telecommunications network to be varied. The latter approach is known as the "scorched node" approach because the locations of nodes are held fixed.[19]

Of course, the scorched node approach, being only a partial optimisation, leads to somewhat higher costs than the scorched earth approach. Despite this, most countries have opted for the scorched node approach.[20] This may be due to reasons of practicality. It is significantly more time-consuming to carry out an optimisation to determine an efficient network from scratch, with few initial constraints. In addition, with more variables subject to discretionary variation the outcome is more likely to be the subject of dispute and legal challenge. Other arguments in favour of the scorched node approach include the following:

1. Australia, in its submission to an OECD roundtable discussion on this topic, argues that "the scorched earth approach would *unduly distort the build-buy* decision of new entrants too far the other way". The FCC makes a similar argument, noting that the scorched node approach "encourages facilities-based competition to the extent that new entrants, by designing more efficient network configurations, are able to provide the service at a lower cost than the incumbent LEC". This argument, has a logical flaw – if the intention is to encourage facilities-based entry, why use an optimising model at all? Why not simply base the regulated prices on the actual costs of the incumbent operator?
2. Germany, in its submission asserts that "it would be *unfair* to request such a degree of optimisation from the incumbent". This last argument can be given a more rigorous economic interpretation, as the next section explains.

In fact, as argued above, the use of the scorched node approach can be simply justified on the basis that it correctly rewards the incumbent for its historical capital expenditure decisions without the need for exceptional depreciation items. (In this case, however, it should be recognised that the cost model is not estimating the least cost of a modern equivalent asset).

5.5. *Accounting for assets earmarked for future demand*

In the previous section we noted that where it was not efficient for the regulated firm to build a network to provide both services A and B today, it should be compensated an amount which reflects the full costs of the original construction of network to provide A plus the full costs of the upgrade to provide service B. But what if it was efficient for the regulated firm to build the additional facilities or capacity at the outset? If the firm is always rewarded for the upgrade costs it will not have any incentive to adequately forecast and provision for future demand, so total costs will be higher. How, then should the regulator handle assets which are built in anticipation of future demand?

If both the regulator and the regulated firm agree that the extra capacity will be required (and that it is efficient to build the extra capacity in advance), the regulated firm could simply be allowed to include the extra capacity in its regulatory asset base.

But what if the regulator is unsure whether the extra capacity will be needed or not, or if it was efficient to build the capacity in advance? For example, the ACCC has a policy that "Investment undertaken in anticipation of future demand is incorporated within the capital base only when the demand materialises".[21]

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

One of the reason's behind this policy is the concern that current customers not be required to pay for capacity which is not required today (and which may not be required tomorrow). The ACCC notes: "While it may be efficient for Telstra to install extra pairs [of copper wire] today in anticipation of future demand, these extra costs should be recovered as the future demand eventuates (otherwise current customers will be paying for services used by future generations)."[22]

In this case the regulator could allow into the regulatory asset base an amount consistent with reasonably forecastable demand and then subsequently adjust the asset base retrospectively if the extra demand actually materialises. The question then becomes how much the regulatory asset base should be adjusted when these "previously excluded" assets are brought back into the asset base.

Consider a model in which there are just two periods. Suppose that there is demand sufficient to justify an investment of I^L today and there is a probability p that tomorrow demand will be higher, requiring an investment of I^H, otherwise demand will remain at the same level. The regulated firm has a choice of making an investment today capable of handling both the demand today and the possible higher demand tomorrow (*i.e.*, an investment of I^H) or of investing for the current level of demand (I^L) today and then, if demand turns out to be higher tomorrow, investing an additional ΔI to provide the required additional capacity. Investing in spare capacity today makes sense if and only if it yields a lower expected present value of investment, *i.e.*, if

$$I^H < I^L + \frac{p\Delta I}{(1+r)} \quad \ldots(6)$$

Suppose that the regulator is of the opinion that customers should not pay for the extra capacity until the extra demand materialises. If the extra demand does not materialise the capital expenditure required is I^L. Now if the extra demand materialises in the second period the asset base must be adjusted so that the expected present value overall is I^H. This means that if the extra demand materialises the asset base in the second period must increase by: $(I^H - I^L)(1 + r)/p$. In other words, if the extra demand materialises the customers in the second period should pay for the extra capacity scaled up to reflect the fact that the capacity was installed one period earlier and to reflect the probability that the capacity would not, in fact, turn out to be needed.

Is there a risk that the customers in the second period will be worse off than if the regulated firm had not purchased the additional capacity at the outset, but simply added capacity as it was needed? The answer is No. If the regulated firm simply added capacity as it was needed, the extra earnings if the capacity was needed would be simply ΔI, but by the assumption that it made sense to invest at the outset rather than wait until the capacity was needed (equation 6) we know that $\Delta I > (I^H - I^L)(1 + r)/p$. In other words, even if the costs of the extra capacity are concentrated in the period when the extra capacity is used, end-users are still better off than if the regulated firm had not provisioned for the extra capacity at the outset.

6. The optimal path of earnings

Up to this point in this chapter we have seen how the regulator chooses a path of earnings which, over the life of the firm, has a present value equal to the firm's capital expenditure. We have also seen how, in practice, this basic task is commonly transformed into the equivalent task of finding a path of depreciation or a path of the asset base. These different ways of viewing the basic task of the regulator are basically equivalent. There is typically an infinite number of different paths of earnings (or depreciation, or of the asset base) which satisfy the basic constraint that the present value of the firm's earnings must be equal to the present value of the firm's capital expenditure.

Amongst the different paths of earnings that satisfy the basic constraint, is there such a thing as an "optimal" path of earnings? In order to answer this question we need to define what we mean by "optimal".

We have already seen one criteria that we can use to select from among the many paths which satisfy the basic constraint – the regulator may wish to choose a path which avoids "rate shock", that is, to choose a path of earnings which is continuous.[23]

If the replacement cost of the underlying asset is continuous, it might be thought that a simple approach would be to base the path of earnings (rather than the path of the asset base as in the previous section) directly on the replacement cost. For example, the earnings in each period could be some constant fraction of the replacement cost. However, this approach requires knowledge of the path of the replacement cost over the entire life of the firm.[24] The constraint that the earnings path be continuous is, in fact, a relatively weak constraint. Many different paths of earnings are both continuous and consistent with the basic constraint.

A better approach is to select a path of earnings which has economically desirable properties. We will explore two economically-based approaches. The first approach looks for an earnings path that does not create an incentive for inefficient duplication of the regulated firm's assets. The second selects the earnings path that leads to the most efficient allocation of resources.

6.1. Preventing inefficient duplication

Suppose that the replacement cost of the firm's assets is declining over time. If the earnings of the regulated firm are not also declining over time there is a risk that the replacement cost of the underlying asset will drop to the point where rival firms or downstream customers find it profitable to duplicate the asset rather than purchase from the regulated firm.[25]

Exactly when a firm will find it profitable to construct or purchase a substitute facility depends on a number of factors which are normally summarised as the "barriers to entry". These include, for example, the extent of economies of scale, switching costs, size and size distribution of end-users and so on. The potential for entry will also depend on the structure of regulated prices and the presence/absence of an access regime (entry may be easier if the entrant does not need to replicate the incumbent's services from the beginning but may simply purchase and resell any services that it does not provide itself).

In the most unrealistic and extreme case of no barriers to entry or exit, in which the threat of entry is the most severe, a rival firm or downstream customer will find it profitable to duplicate the regulated firm's assets at some point in time if, looking forward from that point, the present value of future earnings exceeds the present value of the costs of a brand-new replacement asset. If we are to prevent inefficient duplication in this extreme world, therefore, it must be that in each period the present value of the future stream of earnings must not exceed the replacement cost of the existing asset. In the case where the asset will not be replaced, this is equivalent to the statement that the *level of the asset base must not exceed the replacement cost of the asset* at any point in the future. In particular, if K_t represents the level of the asset base in period t and R_t is the cost of purchasing or constructing an brand new equivalent asset, then it must be that $K_t \leq R_t$ for all time periods.[26]

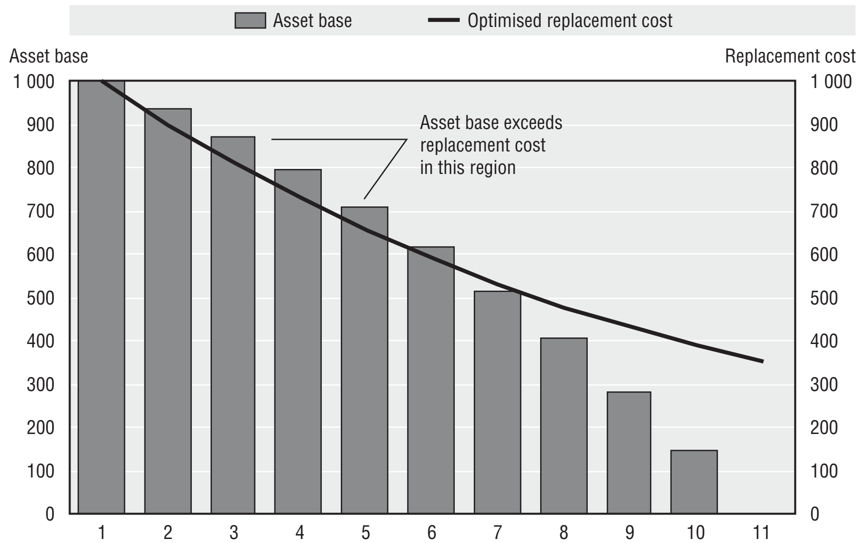
Figure 25 illustrates how a declining path of replacement cost might fall below the level of the asset base (and therefore a risk of inefficient entry) in the case where the regulator sets a constant regulated price. In this case the level of earnings of the regulated firm are given by the value of the simple annuity. As can be seen, when the regulator uses a simple annuity and the replacement cost is falling at a constant geometric rate, the replacement cost may fall below the level of the asset base, leading to a risk of inefficient entry.

In fact it is not entirely accurate to compare the current asset base with the cost of purchasing a brand new asset – a brand new asset has a longer life than the incumbent asset, so a firm considering replicating the incumbent asset would compare not just the present value of earnings over the remaining life of the firm, but also the future earnings it would be saving at the start of the next generation of assets.

If we assume the replacement cost is declining at a constant rate (*i.e.*, $R_t = (1 + \alpha)R_{t-1} = (1 + \alpha)^{t-1}I$) then, the highest possible path for the asset base corresponds to a path of earnings which are also increase at the same rate. In other words, the earnings are given by a fixed percentage of the replacement cost, $E_t = \phi_h R_t$ where:

$$\phi_N = (r - \alpha)\frac{(1 + r)^N}{(1 + r)^N - (1 - \alpha)^N} \qquad \ldots(5)$$

Figure 25. **Asset base exceeds declining replacement cost for a simple annuity**



Source: OECD.

To give some idea of the effect of this approach, suppose that the price of the asset is initially $1000 and declines at 5% per year. The cost of capital is 10% per year. In order to prevent inefficient duplication, while ensuring that the regulated firm recovers its investment cost, the earnings path must be as follows: $195.02, $185.27, $176.00, 167.20 … and so on. In contrast, if the regulator set a simple constant earnings path the earnings in each period would be just $162.75. The need to reflect the falling price of capital goods requires that the price be raised in the first period by about (in this case) 20%.

6.2. Inter-temporal Ramsey pricing

Let's put to one side for a moment the possibility of inefficient duplication of the regulated firm's assets and let's return to the question of whether there is an "optimal" path of earnings. An "optimal" path of earnings is one that leads to an efficient allocation of resources – *i.e.*, the path that maximises total welfare subject to the condition that the regulated firm is able to recover all the costs of its investment.

This is one way of stating the familiar problem posed by Ramsey – how do we allocate a fixed cost over different services (in this case different periods) in such a way as to maximise consumer welfare? The answer to this problem is well known – as long as the price must be raised above marginal cost in some periods to cover the total cost, the margin between the price and marginal cost should be higher in those periods with more inelastic demand.This was discussed in detail in Chapter 1 of this report. A derivation of the Ramsey formula in this context is set out in the appendix.[27]

There are very many implications of this basic insight. Inter-temporal Ramsey pricing might imply, for example, setting low prices in the first period after the introduction of a new service, when demand for the product is weakest. Prices might increase in later periods as consumers learn about the benefits of the service and demand strengthens. Eventually, prices might again decline as the novelty of the new product wears off and rival firms develop substitute products.

Inter-temporal peak-load pricing

When the regulated firm has a limited capacity, the optimal path of prices is likely to involve *a)* lower prices (near marginal cost) in periods when demand is sufficiently low that the capacity of the asset at those prices is not exhausted and *b)* pricing well above marginal cost in those periods when the demand is sufficient to exhaust the capacity of the regulated firm. This approach is called "peak-load pricing".

Baumol (1971) shows that if revenue is sufficient to just cover operating costs in off-peak periods and rises to a level sufficient to ration the capacity in peak periods, and if there are constant returns to scale in new investment, the price can be set equal to marginal cost in off-peak periods and the resulting earnings in peak periods will just cover the full cost of the investment.

Note that the path of prices arising from inter-temporal peak-load pricing might be exactly the opposite of the path of prices that might arise from the use of straight-line depreciation or depreciated replacement cost. For example, consider the case of a firm facing increasing demand and suppose for simplicity that the cost of capital is zero. Under simple straight-line depreciation the earnings of the firm are constant over time. But, if demand is increasing, constant earnings means that prices must be declining as the quantity sold increases. So the simple application of straight-line depreciation would have prices falling over time until a new facility is built, and then increasing sharply when new capacity comes on line.

Newbery (1997) emphasises that such a path of prices is precisely the opposite of what is required for efficiency when the facility is capacity constrained. In contrast to declining prices, the optimal prices would be low at the outset, and increasing over time to efficiently ration access to the facility when the facility becomes congested, and reduced again when the facility is expanded to alleviate the capacity constraint.

Two-part pricing

As is well known, when prices must be raised above marginal cost in order to cover fixed costs, it is often possible to improve total welfare by relying on various forms of price discrimination. One approach is to use two-part pricing. That is, end-users could be charged a two-part tariff consisting of a fixed part and a variable part. If the elasticity of the fixed part of the tariff is low enough, welfare can be improved by raising the fixed component of the tariff and lowering the variable component to the point where the price at the margin is close to marginal cost.

Another possibility is to segregate the market based on classes of customers. The price could then be raised on those customers with inelastic demand and lowered on other customers with more elastic demand.

The use of an inter-temporal price-cap

This broad range of possibilities highlights the fact that, if the regulator is to set prices efficiently, the regulator must have detailed knowledge of *i)* how different groups of consumers respond to changes in the prices for different quantities of the each of the different services sold and *ii)* how that demand responds to changes in the prices of the same and other services in other periods. In practice, the regulator is not likely to have such detailed information.

As we saw in Chapter 1 of this report, where there is a possibility that the regulated firm will have better knowledge of demand than the regulator, it often makes sense to delegate to the regulated firm responsibility for setting individual prices subject to a cap on the weighted average of prices overall.

This result, which is conventionally applied to services provided in a single period, may also apply across periods. If the regulator were able to commit to a contract equal in length to the life of the regulated firm, the regulator could simply place a cap on the weighted average of prices over time. The regulated firm could then choose to set the optimal profile of prices over time – setting lower prices when demand is weaker and higher prices when demand is stronger, and would choose to use various forms of price discrimination, as required by the Ramsey formula. The price cap would ensure that, over time, prices were not so high as to allow the regulated firm to earn excess returns. A derivation of this result is set out in the appendix.

An argument against the use of an inter-temporal price-cap relates to the possibility for strategic behaviour. Specifically, the regulated firm might choose to set high prices in the initial periods of the cap, recognising that the price-cap, if enforced, would require low or even loss-making prices in later periods. In effect, the regulated firm is betting that the regulator will never allow the firm to go bankrupt and will, instead, allow some relief from the price cap in later periods. If the regulator cannot for political reasons allow the regulated firm to become bankrupt or to cease operations, the regulator cannot commit to enforcing an inter-temporal price cap. Perhaps the inter-temporal price-cap could be combined with a price ceiling, to prevent the regulated firm from earning such high returns in early years as to force it into bankruptcy in later years.

It is sometimes noted that the optimal weights in the price cap are proportional to the quantities purchased at the efficient prices, so that the total weighted average of prices is proportional to total revenue. However, it is not the case that a cap on the weighted average of prices is equivalent to a revenue cap. Under a revenue cap the regulated firm has no incentive to raise its total revenue in a way which is optimal for consumers – the regulated firm is indifferent, for example, between collecting all its revenue in one period or spreading that revenue over many periods. Alternatively a firm subject to a revenue cap could charge the monopoly price for the first part of its regulatory period and then (having met its revenue cap) simply shut down in the second part of the regulatory period. This outcome is unlikely to be desirable from the perspective of the regulator.

Regulators sometimes dismiss the suggestion that prices should be differentiated according to different demand conditions (as is required with Ramsey pricing) on the grounds that it is "too hard". It is possible to make several points in response:

- First, price discrimination of various kinds is often a fundamental feature of prices at the retail level. Telecommunications companies are experts at offering a wide menu of calling plans to different classes of customers, with different fixed/variable components and with different prices for calls at different times of the day and for different services. In fact, telecommunications companies are sometimes criticised for engaging in this practice to the extent that consumers are left bewildered by the variety of tariffs on offer. If this discrimination is possible at the retail level is it impossibly difficult at the access level? Indeed, as Chapter 1 of this report showed it may be essential to reflect differentiation at the retail level in differentiation at the access level.
- Second, even here full Ramsey pricing is not feasible, it may be possible to move in the direction of greater forms of differentiation that improve overall welfare. Indeed if existing prices are above marginal cost it is always possible to introduce a new tariff (or to modify an existing tariff) which improves the welfare of both consumers and the regulated firm.
- Third, neglecting demand-side factors entirely when setting regulated prices has one important consequence – it can prevent the firm from carrying-out investment projects which are socially valuable. This point is

clearest for those projects which are marginal in the sense that the cost of the project is large relative to the maximum possible return that the firm can earn from the project. The maximum possible return that a firm can earn involves careful price discrimination exactly along the lines of Ramsey pricing. If the regulator prevents the firm from discriminating in this way the firm may not be able to earn sufficient revenue to justify investing in the project. The more marginal a project, the more important it is that the regulator set regulated tariffs close to the Ramsey optimal prices.

7. The effect of uncertainty

Up until this point we have been discussing a world with no uncertainty. In practice uncertainty over future outcomes is ubiquitous and has a fundamental impact on investment decisions. This, in turn, affects the level of compensation required by the regulated firm to induce it to invest.

The uncertainty may take many forms, such as uncertainty over future demand (including the possibility of the development of substitutes), over future costs, over asset life, over future interest rates and so on. In fact, all the elements of the problem that have we have taken as fixed up until now may be subject to important uncertainty.

We will focus on the case where all of the uncertainty is in the future earnings. That is, the future earnings are unknown but the firm's capital expenditure, asset life and the (risk-free) cost of capital are all known.

In this section we will assume that the uncertainty in earnings is *stationary* – that is, the uncertainty in earnings each period is independent and identically distributed. The effect of this assumption is that every period "looks the same" at the beginning of the period before the earnings have been realised, even though the outcome may be different.

In each period, the regulated prices affect the uncertain earnings of the firm – higher prices will presumably lead to higher earnings, on average, and *vice versa* . Note that the prices set by the regulator could affect not just the expected *level* of the earnings but also their variance and various other characteristics (such as their correlation with other returns in the economy).

The task of the regulator is to set the prices in each period in such a way that, given the nature of the uncertainty in the earnings, investors are just indifferent between keeping their capital in the firm or withdrawing it and investing elsewhere. If we assume that the level of debt of the firm is sufficiently small that the probability of default is negligible, despite the uncertainty in total earnings, the effect of the uncertainty is concentrated on the return on equity. The regulator's problem is then to determine the appropriate return on equity.

 PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

The difficult issue here is determining, given the uncertainty in earnings, an expected level of earnings (*i.e.*, an expected return on capital) which is just sufficient to induce the investors to keep their capital in the firm. Calculation of this required expected return requires the use of financial models which yield, for a given form of uncertainty the expected return on equity that investors require to be induced to hold the underlying asset.

7.1. The capital-asset pricing model

Uncertainty in the external environment, combined with the nature of the regulation on the regulated firm will determine the mean and variance of the earnings of the regulated firm. How should this mean and variance affect the firm's required cost of capital. One of the most well-known models which yields a firm's required cost of capital given the variability in the earnings of the regulated firm, is the capital asset pricing model (or "CAPM").

The CAPM model shows that the returns that investors require for holding a risky asset depends not only on the absolute riskiness of the returns on the asset itself, but on the correlation of the return of that asset with the overall portfolio of assets already held by investors. The portfolio of assets already held by investors is known as the "market portfolio". If we assume that the assets of the regulated firm are small relative to the market portfolio, it is possible to show (see the appendix) that the required *expected* or *average* return of the regulated firm's equity is given by:

$$r_f + \beta\,(ER_M - r_f) \qquad \ldots(7)$$

where:

r_f is the "risk free" rate of return;

β is the correlation of the regulated firm's returns with the market portfolio;[28]

ER_M is the expected return on the market portfolio;[29]

In the case where the uncertainty is stationary (that is, when the uncertainty in earnings is independently and identically distributed) we can apply the results which we derived in the previous sections, with the proviso that the discount rate is not the risk-free rate *r* but the required expected return on equity given by the CAPM formula, equation 7.

All of the results of the previous sections still apply, provided that it is understood that the regulator is not controlling the path of actual earnings (which are random) but the average or expected earnings in each period given the prices set by the regulator. In each period, the regulator chooses a path of expected earnings (or of depreciation or of the asset base) so that the expected earnings are given by the sum of the return on capital (at the required cost of capital) and the return of capital (given by depreciation). The basic task of the regulator can be re-expressed as the choice of a path of expected earnings (rather than actual earnings) so that the net present value of the expected net payment to investors (discounted using the required cost of capital) is zero.

Note that the fundamental cost allocation problem remains. The precise level of expected earnings in any one period remains discretionary. As before, we must add additional constraints to the problem if we are to obtain an allocation of costs with desirable economic properties.

7.2. Criticism of the CAPM and other approaches

The use of the CAPM as a tool for estimating the required return on equity is not universally accepted amongst economists. There are at least three strands of criticisms:

1. First, the use of the CAPM model as a tool for regulation requires that the firm's "beta" is stable over time and independent of the actions of the regulator. Neither assumption is likely to be true. A firm's beta is likely to change with market conditions – so measures of a firm's beta based on historic data will be inaccurate. This criticism applies not just to the CAPM but to all cost of capital models based on historic data.[30]

   More importantly, beta will typically depend on the *level and structure of the regulated prices*. Suppose that the regulated prices of the firm are allowed to vary with key input costs (such as the cost of energy sources for an electricity generator). In this case the regulated firm will be insulated from the risk of changes in the price of key inputs, lowering its measured beta. Alternatively, if the retail prices of the regulated firm take the form of a two-part tariff, with the bulk of the revenue coming from the "fixed" component of these charges, the regulated firm will be largely insulated from the risk of changes in demand, again lowering its beta.[31]

   Since beta depends on the regulation and the regulation depends on beta, the best that the regulator can do is to seek a fixed-point solution – *i.e.*, a value of beta which yields a degree of variability in the firm's earnings consistent with the original choice of beta. Ideally this fixed point solution would be stable – small variations in the measured beta would lead to smaller deviations from the fixed point in the following period. However if lower prices are associated with greater variability in earnings (and therefore a higher beta), there may be *no* stable fixed point solution.

2. Second, it is sometimes argued that the regulator chooses an inappropriate beta – *e.g.*, the particular beta chosen would be appropriate for a firm which is not subject to regulation while regulation has the effect of changing the distribution of returns of the regulated firm. One of the possible effects of regulation is to truncate the distribution of returns – the regulator steps in

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

to limit any upside gains but does not symmetrically restrict downside losses of the regulated firm. Ergas (2001) writes:

"Regulators may … find it difficult to allow regulated firms to secure what appear to be supra-normal earnings from an essential facility – even when these come as a return for a socially desirable action, such as efficient investment or superior business skill. To the extent this is the case, the regulator will expropriate 'excess profits' associated with favourable states of the world. Conversely, when the regulated firm suffers from unfavourable conditions, it has to bear the full consequences. Regulation of this kind can be seen as involving a game of 'heads you lose, tails you break even'."[32]

In fact, the regulator may not need to actively step in for truncated returns to arise – if the regulator sets a price ceiling but not a price floor and the uncertainty is in demand then the regulated firm may choose to price up to the price cap (*i.e.*, the price cap may bind) when demand is high but not when demand is low. Even if the uncertainty in demand is normally distributed, the resulting uncertainty in earnings will exhibit a truncated distribution.

For smaller variations in the level of demand, the price cap may bind for all levels of demand and the only uncertainty faced by the firm is in the quantity purchased. In this case, provided that *a)* the regulator accurately sets the price cap equal to the expected quantity purchased (given that regulated price and possible demand levels) and *b)* the marginal cost of the firm is constant over the relevant range, the regulated firm will, on average, earn no more or less than the expected profit.[33]

3. Third, there may be better financial models for estimating the required cost of capital. The CAPM relates the required return to a single independent variable – the risk premium on the "market" portfolio. Other financial models, which relate the required return to more independent variables may be able to do a better job at predicting the required return. For example, it may be possible to obtain a better estimated of the required return on capital by extending the CAPM to incorporate a term which reflects this skewness.[34] One such model is the "three-moment model".[35] One study estimates the required cost of equity for 60 utilities in the US in the period 1976-1980.[36] This study finds that the three-moment cost of equity is around 1.25% higher than the required cost of equity given by the CAPM.

   Another approach is to abandon the CAPM entirely and to use a more sophisticated financial model to determine the required return, such as the so-called "Arbitrage Pricing Theory" or APT. This approach estimates the required return on equity as a linear combination of a number of "factors" which are determined using statistical techniques. Bower, Bower and Logue (1984) use this technique to estimate the returns on gas and electricity utilities in the US in the period 1971-1979. They find an APT cost of equity 2.3% lower than CAPM for electricity utilities and 1.9% higher than CAPM for gas distribution utilities.[37]

In conclusion, although CAPM is widely used, its shortcomings should be recognised. In particular, the asset beta is likely to vary according to market conditions and the decisions of the regulator. There is no guarantee that an attempt to set beta on the basis of historic information will lead to an accurate estimate of beta or will converge to an equilibrium value.[38]

7.3. Accounting for the possibility of technological obsolescence

The previous section assumed a particularly simple form of uncertainty – that the uncertainty in earnings was stationary. Therefore, despite the uncertainty, each period "looks the same" as every other period. No new information is learned as time progresses.

However, this is not necessarily the case. It might be that the level of demand in one period is affected by the level of demand in previous periods. If demand happens to be high in earlier periods, it may be more likely to be high (or to grow at a faster rate) in future. Alternatively, it may be that demand for the services of the regulated firm are affected by technological developments. If a new competing service is developed, it may be that there is a large and permanent drop in the demand for the services of the regulated firm. Incorporating this form of uncertainty requires certain adjustments to our basic model.

With non-stationary uncertainty, the previous building block model can still be applied provided that *a)* at the beginning of each period the regulator accurately re-values the asset base to be equal to the present discounted value of future earnings in the light of any new information and *b)* the asset base at the end of any period (or equivalently, the depreciation in any one period) is equal to the *average* or *expected* value of the re-valued asset base at the beginning of the next period.

Let's explore this result in the context of a simple example in which the regulated firm has an asset with an infinite life but each period there is a constant probability p that earnings will drop to a low level and remain there forever. This might be due, for example, to the development of a rival technology that provides a superior service. It might also be due to a regulatory decision to reduce the size of the asset base to reflect a lower replacement cost of the asset.

This example is developed more fully in the appendix to this chapter. There we show that the allowed earnings of the firm can, as before, be expressed as the sum of the return on capital and the return of capital. The difference is that here the return of capital (*i.e.* the depreciation) reflects the

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

expected change in the value of the asset base, which may not reflect the actual change in the asset base as the asset base is re-valued every period.

In other words, it is possible to use a variation of the basic building block model, provided certain adjustments are made. Problems arise, however, when the regulator attempts to use the standard building block model and does not re-value the asset base in the light of new information (in this case the information that in the present period the regulated asset has not yet become obsolete). If the regulator attempts to use the standard building block model the firm will be systematically under-compensated. The reason is that under the standard building block model, the total depreciation allowed to the regulated firm cannot exceed the original cost of the asset. Yet, unless the firm is allowed to depreciate 100% in the first period, at any point in time the firm faces the possibility that demand will drop to zero before it can depreciate any further. On average the firm will not be able to recover the full costs of its investment. In other words, in this context there is no path of depreciation (short of depreciation 100% in the first period) which ensures the firm will be able to earn its required rate of return on average.[39]

If the regulator insists on the simple building block approach (with no asset re-valuation) then some additional compensation must be offered to the regulated firm or the regulated firm will fail to invest. For example, the firm could be granted a higher rate of return or a higher regulatory asset base at the outset.[40] In these cases the extent to which the allowed rate of return or allowed asset base needs to be raised depends on the path of earnings/ depreciation. If the chosen path defers most depreciation until later in the life of the asset, the risk borne by the firm is larger and the firm will require a larger initial asset base. On the other hand, if most of the depreciation occurs early in the life of the asset the risk borne by the firm is limited and the size of the increase in the initial asset base is smaller. In the extreme case when the firm is allowed to depreciate the entire cost of the asset in the first period, it will not need an increase in the asset base at all.

These adjustments are rather *ad hoc* and may create further problems of their own. A more consistent approach, where it is feasible, is, as already mentioned above, to systematically re-value the asset base in the light of new information and to ensure that the end-of-period asset base is equal to the expected value of the beginning-of-period asset base.

Note that, of course, it is possible that a firm facing an investment decision faces not a risk that demand will drop in the future, but the possibility that technological advances will permanently enhance demand for the services of the firm. For example, the invention of the fax machine increased demand for telephone services. Where there is a risk of technological enhancement of this kind the regulated firm should be compensated less in the early periods (before the technological enhancement occurs) rather than more as in the discussion above.

To make this concrete, consider the following example. Suppose that a firm invests in an infinitely-lived asset. Since this asset is infinitely lived there is no depreciation – earnings are given by the "return on capital" alone. Demand is initially "low" but each period there is a 5% chance that demand will shift to "high". Once demand shifts to high it stays there forever. The optimal level of earnings per period in the high state is assumed to be $120. The cost of capital is 10%. The cost of the initial investment is $1 000. Given these parameters, if the firm is to break even the earnings in the low state must be $90. In other words, the earnings of the firm in the low state is *less* than the "return on capital" (which would be $100). The firm is still adequately compensated overall, because of the 5% chance that it's earnings will increase from $90 to $120.

7.4. Uncertain future demand

In the previous example the regulated firm invested knowing that demand was sufficient today but facing a risk that demand would disappear in the future. In the case of "greenfields" investment a regulated firm may expect that demand will be adequate on average but may be uncertain about the actual level of demand that will materialise in the future. As in the previous example, if the regulator applies the simple building-block model the regulated firm will fail to invest, as the following example shows:

Suppose that the regulated firm must make an investment in an asset which costs $1 000 and lasts just a certain finite life. Suppose that future demand is uncertain. Demand in the future could be permanently high or permanently low. Suppose that when demand is low the firm cannot recover sufficient revenue to cover its operating costs, so that it is better to shut down. The firm must invest before it is able to learn whether demand is high or low. Suppose, for simplicity that the firm makes no earnings in the first period.

As before, we can apply the modified version of the building block model to this problem. We can imagine that the firm is re-valued at the beginning of the second period to reflect the new information that is available about demand. Suppose that the regulator sets the earnings if demand turns out to be high to be such that the firm just expects to earn a normal rate of return (in other words, the firm expects to earn $1 000 in present value terms on average). Since there is a probability that, at the end of the first period, the firm will be re-valued downwards if demand turns out to be low, if the firm is to make a normal rate of return it must be that, if demand turns out to be high, the firm is re-valued upwards (*i.e.*, to a value higher than $1 000).

 PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

As before, problems arise when the regulator applies the traditional building block model without adjustment. Suppose, in particular, that the regulator does not re-value the firm at the start of the second period. Then the most that the firm can be worth at the start of the second period is just $1 000. But then, at the start of the first period, since there is a probability that demand will turn out to be low (and the firm worthless), the firm can only be worth less than $1 000 on average. In other words, if the regulator does not re-value the assets of the firm, and instead applies the simple building block model, this project will not be undertaken.

This result has been expressed by Jerry Hausman as follows:

> "Cost based regulated pricing is not appropriate in the presence of considerable demand uncertainty and rapid technological change … because under such regulated outcomes investments which succeed at best earn back their cost, while investments which do not succeed earn back less than their costs. Hence cost-based regulation of only those investments of a firm that ex-post turn out to be successful will lower the firm's overall ex-ante returns. Investment levels will decrease below economically efficient levels."[41]

If the regulator does not re-value the assets of the regulated firm, the regulator must increase the earnings of the firm above the level given by the simple building block model. This might be difficult for the regulator as the regulated firm must be compensated for bearing a risk in the past, *even when that risk did not materialise*. The regulator must not only ensure that the regulated firm is compensated for the stationary and non-stationary risks that it bears in the current period, but also for certain risks that the firm bore in the past, even when the potential negative outcomes did not in fact come to pass.

The key conclusion here is that the simple building block model is inappropriate in a context in which new information is revealed over time. In this context, it is necessary to use a variation of the building block model in which, each period, the value of the firm is re-calculated on the basis of any new information. The end-of-period asset base (and therefore the depreciation) should reflect the expected value of any re-valuation.[42]

7.5. Discussion

Firms that are regulated, or potentially regulated, often express the concern that regulation will reduce incentives for new investment. Does the preceding analysis shed some light on this debate?

There is perhaps a conventional wisdom that regulation only deters investment in the presence of uncertainty. This is not strictly true. Even in world where future outcomes are certain, regulation can decrease the incentives for new investment. The reason is that regulation, by limiting the extent to which the regulated firm can adjust its prices to suit demand conditions, limits the potential earnings of the regulated firm and therefore can deter valuable investment. For example if a regulator sets a constant price in the face of increasing demand, the regulated firm may not be able to recover enough earnings to justify the investment even though the investment is socially valuable. Indeed, as argued earlier, the more marginal the investment, the more important it is for the regulator to choose efficient price structures, or the investment project will not be undertaken.

In the case of those projects which are not at all marginal (*i.e.*, where there is a wide margin between the maximum potential earnings of the firm and the capital costs), the regulator has wider discretion to set prices which, although harmful to overall welfare, do not deter investment. As we have seen throughout this chapter, provided the regulator follows a set of rules (such as those given by the building block model) and provided the regulator is able to correctly measure the cost of capital, the regulator can ensure that the regulated firm will always be fully compensated for its investment. Indeed, since the regulated firm will always be fully compensated for its investment, whatever the level of that investment, it is usually argued that such firms will have an incentive to over- rather than under-invest.

Where there is uncertainty in future demand the regulator should (from a theoretical perspective) set tariffs that are appropriate in the light of all available information about the current level of demand and the future path of demand. The standard building block model must be adjusted to reflect this new environment of uncertainty. In particular, in this context, the regulatory asset base should be re-valued each period in the light of any new information about the appropriate future path of tariffs.

We have shown that if the regulator applies the simple building block model in this context, investment may be deterred. The examples above showed that, in a situation where the regulated firm faced a constant probability of technological obsolescence, the level of depreciation should be higher than in the simple building block model. In the second example, where there was a possibility that future demand would turn out to be too low to justify investment, the regulated firm must be allowed earnings consistent with a higher asset base if the firm is to earn a normal rate of return on average.

The size of this investment-deterring effect will depend on factors such as the magnitude of the uncertainty and the nature of the investment projects that are being considered. In the case of telecommunications it seems likely that the uncertainty about future demand is significantly greater for entirely new services (such as the so-called "3G" services) than for conventional voice services over the fixed network.

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Optimal price regulation in a world of uncertainty may prove difficult. Forecasting demand and costs more than a short period of time into the future may be close to impossible. Is it possible to accurately periodically re-value the asset base of a regulated firm? Is it possible to estimate in advance the expected value of that re-valuation? If the answer to these questions is no, regulators may be forced to use second-best approaches to handling uncertainty and face a trade-off with reduced incentives for investment.

8. Conclusion

This chapter opened with the observation that many telecommunications regulatory regimes require cost-oriented tariffs. The statement that "regulated prices should be based on costs" is potentially misleading.[43] A large number of different cost allocations (and therefore prices) are consistent with the regulated firm recovering its costs. Among these various cost allocations regulators should select a cost allocation with desirable welfare properties. Unfortunately many regulators simply choose a set of prices based on simple rules of thumb such as straight-line depreciation. The failure to reflect demand conditions (or other factors such as the potential for entry) in regulated prices runs the risk of lowering overall welfare and deterring efficient investment. Loosely speaking the *structure* of prices should be determined by demand-side factors while the *level* of prices should be chosen to ensure cost recovery. Regulated prices *should* be based on costs but they should not be based *only* on costs.

In passing, this chapter highlighted why regulators (including competition authorities) should be circumspect about their ability to detect monopoly prices. Regulators usually attempt to detect monopoly pricing through a comparison of current prices and costs in the current period. But where there are sunk investments such cost estimates inevitably involve an allocation of common costs. The way those costs are allocated obviously has an impact on the measured profits. Detecting excess returns requires a comparison not just of prices and costs in one period, but over the entire life of the regulated firm. Furthermore, in a world of uncertainty detecting excess returns requires a consideration of all possible paths of prices starting from the point at which the investment was made. Obviously the information requirements are formidable.

This chapter focused on some of the implications of the simple building block model. In a world of certainty the simple building block model facilitates the process of choosing a cost allocation which ensures the regulated firm is adequately compensated. However the examples presented here highlight how the application of the simple building block model can deter investment in a context of uncertainty. This chapter has presented a variation of the building block model which relies on periodic re-valuation of the asset base but further work is necessary to determine whether or not this approach is practical. If not regulators may have to make do with imperfect approaches with the possibility of reduced investment that this brings.

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Appendix to Chapter 3

Original cost

It is relatively common for regulators to focus on the choice of the path of depreciation. Some of the more common paths of depreciation that are chosen are as follows:

1. *Uniform or straight-line depreciation*, which sets the depreciation to a constant amount in each period (*i.e.*, D1 = D2 = D3 = … = DN = $\frac{I}{N}$). This approach is by far the simplest and is used by a number of regulatory authorities.

   With uniform depreciation the asset base follows the following path:

   $$K_{t+1} = \frac{(N-t)}{N}I$$

   The corresponding earnings path is declining:

   $$E_t = \frac{I}{N}(r(N-t+1)+1)$$

   The resulting earnings has a jump when the asset is replaced unless the replacement cost of the asset after N years drops to $\frac{I}{(1+rN)}$. This is illustrated in Figure A.1.

2. *Geometric depreciation*, in which the depreciation in each period is a certain fraction 1 + α of the depreciation in the previous period, (*i.e.*, $D_t = (1+\alpha)D_{t-1}$ which gives: $D_t = 1\frac{\alpha(1+\alpha)^{t-1}}{(1+\alpha)^N-1}$).

   The corresponding asset base follows the following path:

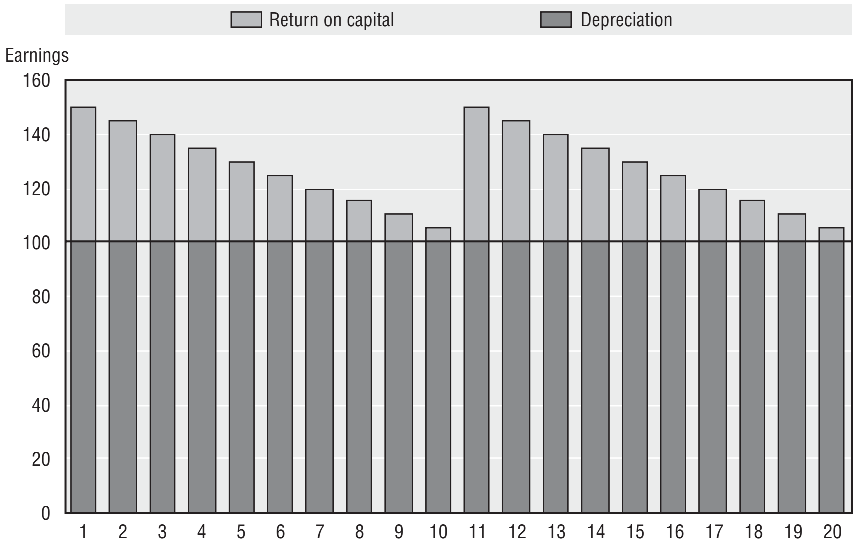
   $$K_t = I\frac{(1+\alpha)^N-(1+\alpha)^{t-1}}{(1+\alpha)^N-1}$$

   and the corresponding earnings path:

   $$E_t = rI\frac{(1+\alpha)^N+(1-\alpha/r)(1+\alpha)^{t-1}}{(I+\alpha)^N-1}$$

Note that when the depreciation is just increasing at the rate of interest (*i.e.*, α = *r*) the resulting path of earnings is constant. The resulting earnings has a jump when the asset is replaced unless the replacement cost of the asset after N years drops to $I\left(\frac{(1+\alpha)^N+(1-\alpha/r)}{(1+\alpha)^N+1-\alpha/r}\right)$, which is just *I* in the case when $\alpha = r$.

Figure A.1. **Uniform depreciation**



Notes: Capital expenditure $1 000 in year 0 and $1 000 in year 10; Asset life 10 years; Interest rate 5%;
Source: OECD.

Figure A.4. illustrates the case of geometric depreciation in the case where the replacement cost of the asset is declining at 5% per year. Note that in all of the first ten periods of the life of the asset the present discounted value of future earnings exceeds the replacement cost, leading to an incentive for duplication of the asset.[44]

1. *Sum-of-the-years'-digits,* in which the depreciation is declining, according to the formula:

$$D_i = I\frac{N-t+1}{\sum_{i=1}^{N} N-i+1} = I\frac{2(N-t+1)}{N(N+1)}$$

The corresponding path for the asset base is given by:

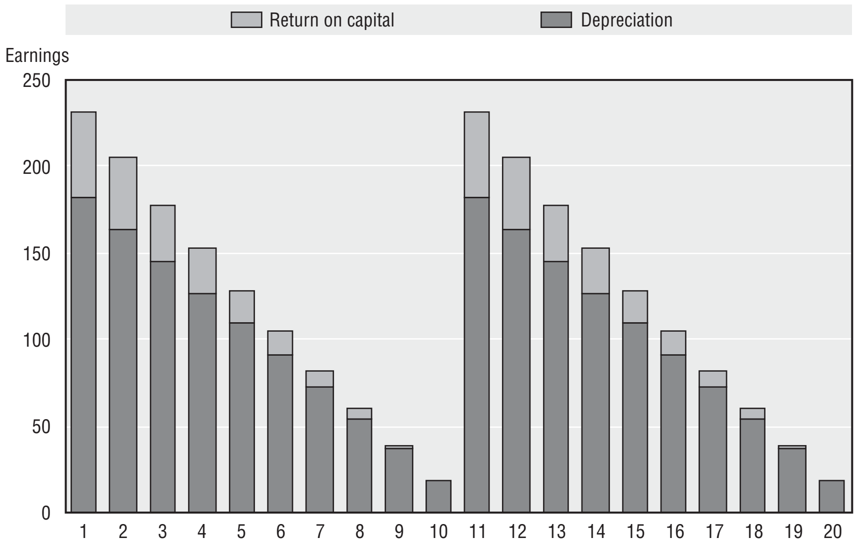
$$K_t = I\frac{(N-t)(N-t+1)}{N(N+1)}$$

and the earnings path:

$$E_t = rI\frac{(N-t+1)(N-t+2/r)}{N(N+1)}$$

This earnings path has a jump when the asset is replaced unless the replacement cost of the asset drops to *zero* at the end of the life of the asset.

Figure A.2. **Sum-of-the-years'-digits depreciation**



Notes: Capital expenditure $1 000 in year 0 and $1 000 in year 10; Asset life 10 years; Interest rate 5%;
Source: OECD.

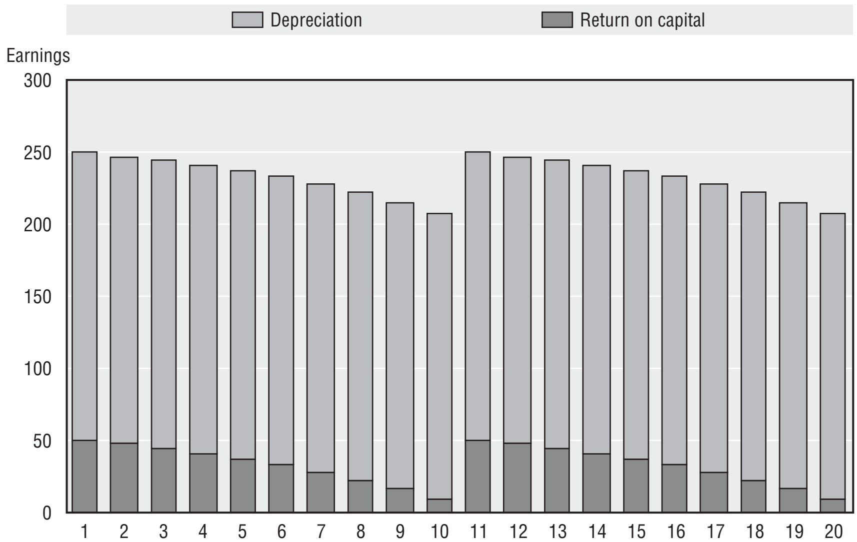
Trended original cost

One variant of the original cost approach involves scaling up the asset base each period by the rate of inflation.[45] When straight line depreciation is applied in this context it can be shown that the outcome is equivalent to the simple original cost approach with depreciation given by:

$$D_t = I(1+i)^{t-1}\frac{[1-i(N-t)]}{N}$$

This approach is illustrated in Figure A.3 for the case where the inflation rate is 15%, the real interest rate is 5% and the replacement cost is constant.

Figure A.3. **Trended original cost**



Notes: Capital expenditure $1 000 in year 0 and $1 000 in year 10; Asset life 10 years; Interest rate 20%; Inflation rate 15%.

Source: OECD.

Depreciated replacement cost

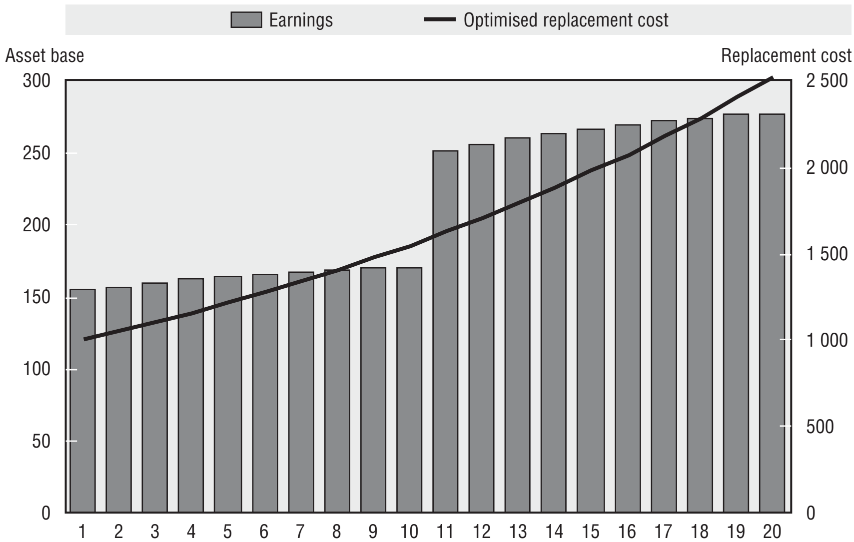
Under the DORC approach, the asset base is a function of the current replacement cost of the underlying asset (*i.e.*, the current cost of a new modern equivalent asset). This replacement cost is then "depreciated" an amount corresponding to the shorter life of the existing asset. The DORC approach can be expressed mathematically as follows. Let R_t be the optimised replacement cost of the asset in period *t*. The DORC approach amounts to defining the path of the asset base $K_1, K_2, \ldots, K_N$ as $K_t = \alpha_t R_t$ where α_t is a path of numbers satisfying $\alpha_1 = 1$ and $\alpha_{N+1} = 0$.

Using mathematical notation, with uniform depreciation the path of capital is given by the expression: $K_t = \alpha_t R_t = \frac{(N-t+1)}{N} R_t$. Using this expression we can work backwards to find the corresponding path of depreciation, which in this case is:

$$D_t = \frac{R_t}{N} - \frac{N-T}{N}(R_{t+1} - R_t)$$

We can see from this expression that as long as the optimised replacement cost is constant, this approach reduces to simple uniform depreciation discussed earlier. However, whenever the optimised replacement cost is not constant the resulting depreciation will not correspond to uniform depreciation.

Figure A.4. **Depreciated replacement cost**



Notes: Capital expenditure $1 000 in year 0 and $1 629 in year 10; Asset life 10 years; Interest rate 10%; Replacement cost increasing at 5 % per annum.

Source: OECD.

Figure A.4. illustrates the path of earnings when the path of the asset base is given by depreciated replacement cost, when the optimised replacement cost is increasing at a steady percentage per year and the replacement cost is depreciated with simple uniform depreciation. Note the discontinuity in earnings at the time when the asset is replaced.

Preventing inefficient entry

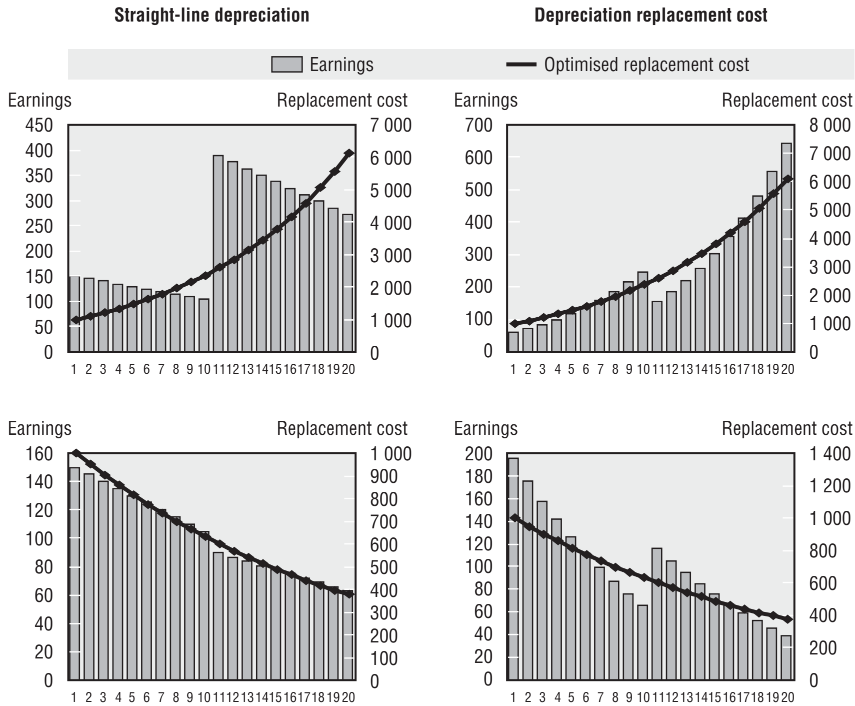
Let's suppose that the regulator sets the path of earnings E_1, E_2, … into the indefinite future. We will assume that the path of earnings has the property that every time the asset needs to be replaced the present value of earnings over the life of the asset equals the replacement cost of the asset, *i.e.*,

$$R_t = \sum_{s=1}^{N} \frac{E_{t+s}}{(1+r)^{t+s}} \text{ for t = 0, } N, 2N, 3N, \ldots$$

Let's define a new expression, $PV_t(s)$, as follows:

$$PV_t(s) = \frac{R_{t+s}}{(1+r)^s} + \frac{R_{t+N+s}}{(1+r)^{N+s}} + \ldots$$

Figure A.5. **Comparison of straight-line depreciation and depreciated replacement cost**



Notes: In the upper two graphs the replacement cost is increasing at 10% per annum, capital expenditure is $1 000 in year 0 and $2 594 in year 10; in the lower two graphs the replacement cost is decreasing at 5% per annum, capital expenditure is $1 000 in year 0 and $598 in year 10. In all cases the asset life is 10 years; interest rate 5%.

Source: OECD.

Using the above result, we have that:

$$PV_t(N-t) = \frac{R_N}{(1+r)^{N-t}} + \frac{R_{2N}}{(1+r)^{2N-t}} + \ldots = \sum_{s=N+1}^{\infty} \frac{E_s}{(1+r)^{s-t}}$$

Let's suppose that there is a threat of perfectly contestable entry – that is, the entrant can enter costlessly by constructing a new asset of its own and stealing the entire market from the incumbent. The net present value investing at time *t* and taking over the business of the incumbent is:

$$NPV_t = \sum_{s=t}^{\infty} \frac{E_s}{(1+r)^{s-t}} - PV_t(0) = \sum_{s=t}^{N} \frac{E_s}{(1+r)^{s-t}} + \sum_{s=N+1}^{\infty} \frac{E_s}{(1+r)^{s-t}} - PV_t(0)$$
$$= K_t + PV_t(N-t) - PV_t(0)$$

Since we want the net present value for the entrant to be non-positive we must have $K_t \le PV_t(0) - PV_t(N-t)$ for t = 1, 2, …, N

Hence the highest possible path of the asset base consistent with deterring entry is given by:

$K_t = PV_t(0) - PV_t(N-t)$ for t = 1, 2, …, N

Ramsey pricing

Suppose that we have a natural monopoly with a long-lived investment. We will seek the set of prices which maximise consumer welfare subject to the constraint that the present value of earnings is sufficient to cover the cost of the initial investment.

Let $W(P_1, P_2, \ldots, P_N)$ be the consumer welfare when the path of prices is $P_1, P_2, \ldots, P_N$. Similarly, let $\Pi(P_1, P_2, \ldots, P_N)$ be the present discounted value of the earnings of the firm when the path of prices is $P_1, P_2, \ldots, P_N$. If we assume that demand in each period is independent of demand in other periods then we can write $W(P_1, P_2, \ldots, P_N) = \sum_{i=1}^{N} \frac{w_i(P_i)}{(1+\delta)^i}$ where $w_i(P_i)$ is the consumer surplus in each period and δ is the rate of time preference. Similarly, $V_1 = \Pi(P_1, P_2, \ldots, P_N) = \sum_{i=1}^{N} \frac{E_i(P_i)}{(1+r)^i}$ where $E_i(P_i)$ is the earnings (profit) earned in period *i* (*i.e.*, revenue less operating costs in that period $\pi_i(P_i) = (P_i - c)Q_i(P_i)$).

Now, if we maximise $W(P_1, P_2, \ldots, P_N)$ subject to $V_1 = \Pi(P_1, P_2, \ldots, P_N) \geq I$ we find that the price in each period must be given by the expression which relates the price-cost mark-up each period to the inverse of the elasticity of demand in that period.

$$\frac{(P_i - c)Q_i' + Q_i}{Q_i} = \frac{\phi^i}{\lambda}$$

which is more well-known in the following form:

$$\frac{P_i - c}{P_i} = \frac{1}{\varepsilon_i}\left(\frac{\phi^i - \lambda}{\lambda}\right)$$

where $\varepsilon_i = \frac{P_i Q_i'(P_i)}{Q_i(P_i)}$ is the elasticity of demand, $\phi = \frac{1+r}{1+\delta}$ is the ratio of the interest rate to the rate of time preference and λ is a constant.

In the special case when the interest rate and the rate of time preference are equal, the price in each period is proportional to the elasticity of demand in that period – in other words, if demand is constant over several periods the price will also be constant. This implies in turn that the quantity consumed is constant and therefore the revenue (and earnings) are constant. As mentioned earlier, this corresponds to a geometric depreciation schedule with the amount of depreciation in each period increasing at a rate corresponding to the interest rate (*i.e.*, $D_2 = (1 + r)D_1$, $D_3 = (1 + r)^2 D_1$, …).

In the case where the demand curve has constant elasticity, the above expression says that the price-cost mark-up should be increasing over time, or decreasing over time depending on whether the rate of time preference is larger than the interest rate. When the rate of time preference is larger than the interest rate, it is preferable to consume now and to put off the exercise of restraint until later – this corresponds to lower prices today, in exchange for higher prices towards the end of the life of the asset. Conversely, when the rate of time preference is lower than the interest rate it is preferable to defer consumption until nearer the end of the life of the asset.

More generally, the demand in each period may depend in a sophisticated way on prices in past or future periods. For example, it might be that there is a "bandwagon effect" where demand in later periods depends in part on consumption in earlier periods. If this is the case it might be preferable to keep prices low in early periods in order to increase consumption in order to be able to maintain higher prices in later periods.

Inter-temporal price cap

Suppose that we allow the regulated firm to choose its prices $P_1, P_2, \ldots, P_N$ subject to a cap of the form $\sum_{i=1}^{N} w_i P_i \leq k$. The firm will choose the prices which maximise its profit $\Pi(P_1, P_2, \ldots, P_N) = \sum_{i=1}^{N} \frac{\pi_i(P_i)}{(1+r)^i}$ subject to this cap. The firm will then choose its prices to satisfy the following expression:

$$\frac{(P_i - c)Q_i' + Q_i}{Q_i} = \mu(1+\delta)^i \phi^i \frac{w_i}{Q_i}$$

So, if the regulator chooses the weights proportional to the quantities in each period discounted by the rate of time preference, the regulated firm will choose the socially-efficient prices.

Technological obsolescence

Consider the following example. Suppose that demand can be in one of two states – high or low. When demand falls to the low state it stays there forever. There is a probability p that demand will drop to low in any period. Suppose that in period 1 demand is in the high state. Let E_t^H and E_t^L be the earnings of the firm in the high and low state of demand in period t.

PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

We can write the expected present discounted value of the firm in the high and low states as:

$$V_t^H = \frac{E_t^H}{(1+r)} + \frac{1}{(1+r)}((1-p)V_{t+1}^H + pV_{t+1}^L)$$

and $$V_t^L = \frac{E_t^L}{(1+r)} + \frac{E_{t+1}^L}{(1+r)}$$

Re-arranging gives us that $E_t^H = rV_t^H + D_t^H$ where $D_t^H = V_t^H - E[V_{t+1}^H]$ is the expected depreciation and $E[V_{t+1}^H] = (1-p)V_{t+1}^H + pV_{t+1}^L$ is the expected

value of the re-valuation at the beginning of period $t + 1$.

There are other ways that the firm could compensated for this additional risk of obsolescence. For example, the firm could be granted a higher rate of return on its capital stock. Specifically, the firm could be allowed the rate of return

$r + \alpha_t$ where $\alpha_t = p\left(1 - \frac{V_t^L}{V_t^H}\right)$. This raises the question of the extent to which the

CAPM already adequately compensates the firm for risks of this kind.

In the case where $V_L = 0$ this reduces to replacing the interest rate r with $r + p$. In other words, if the interest rate is 5% and there is a 2% chance of technological obsolescence each period, the earnings of the firm should be 7% of the firm's asset base.

Option to delay

Suppose that in period 1 demand is zero but can jump to one of two states- high or low from period 2 onwards. The value of the firm in period 1 is just:

$$V_1 = \frac{1}{(1+r)}((1-p)V_2^H + pV_2^L)$$

In order to induce investment today it must be that the value of the firm in the first period is equal to the investment requirement, *i.e.*, $V_1 = I$. But, if the regulator does not allow the regulated firm a higher asset base when demand is high, then $V_2^H \leq I$ and since $V_2^L = 0$ we have that

$V_1 = \frac{1}{(1+r)}((1-p)V_2^H + pV_2^L) \leq \frac{1-p}{1+r}I < I$ so the firm will not invest.

If the firm is to invest, the asset base must be increased to at least $V_2^H \geq \frac{(1+r)}{(1-p)}I$.

Notes

1. The Telecommunications Act Sec. 252 (d) (1) (a) states that regulated prices for interconnection and network elements of ILECs "shall be *i)* based on the cost (determined without reference to a rate-of-return or other rate-based proceeding) of providing the interconnection or network element". ILECs can, however, also enter into bill-and-keep interconnection arrangements.
2. Directive 97/33/EC article 7 states: "Charges for interconnection shall follow the principles of transparency and cost orientation." The Reference Paper of the Negotiating Group on Basic Telecommunications states (article 2.2 (b)) that interconnection is provided at "cost-oriented rates".
3. There is a sense in which the cost allocation problem can only be "solved" by determining the efficient set of final prices and then working backwards to determine the corresponding allocation of costs which is compatible with those prices. In other words, it is often more meaningful to talk about "price-based costs" than "cost-based" prices. See Salinger (1998).
4. In other words, most accounting measures of profitability are worthless. This point is emphasised by Fisher and McGowan (1983). The profitability of a firm can only be assessed through knowledge of the entire path of earnings over the life of the asset. For example, if a firm consists of one asset that lasts five years, it is possible to unambiguously measure the total cost of providing the services of the firm over that five-year period. Given information on revenues over the whole five years it is possible to precisely determine whether or not the firm is making monopoly profits over that period.
5. Here and throughout this chapter we are assuming that the regulator controls *all* the earnings streams of the regulated firm. It is not possible to efficiently regulate a monopolist when the regulator only has direct or indirect control over a subset of the services for which the monopolist has market power. In the special case in which the final services produced by the monopolist are a perfect substitute for the final services produced by entrants and there is a high level of competition in the market for final services, the regulator can dispense with final price regulation on the incumbent as unnecessary (the market power of the incumbent is indirectly disciplined by competition in the market for final services).

   In the case where the regulated firm has other unregulated streams of revenue the regulated firm will only continue to invest in the regulated services if the incremental revenue from the regulated services exceeds the incremental cost. If the unregulated streams of revenue would be put at risk by failure to invest in the regulated service this unregulated revenue should be included within the incremental revenue. If this unregulated revenue is large enough relative to the cost of an investment, the revenue from the regulated tariffs could be correspondingly smaller without affecting the decision whether or not to invest.
6. This is not to imply that these two terms have the same meaning in all countries.

7. This approach is sometimes known as "economic depreciation" after Hotelling (1925), but the term "economic depreciation" seems to be used to refer to so many different practices that this term no longer seems to have a precise meaning.

8. One immediate consequence is that except in special circumstances, there should be no need to include land within the asset base of a regulated utility (although where the land has a special configuration, as would be the case for a railroad, it makes sense to include the additional costs that would be involved in obtain that configuration within the asset base of the regulated firm).

9. Ehrhardt (1994), page 167.

10. If it is useful, it can be assumed that the assets we are discussing have the characteristic of constant productivity over their life – so that there is precisely no physical deterioration of the assets over their useful life. This is the so-called "one-hoss shay" assumption.

11. As one example, Ehrhardt (1994) summarise the procedure as follows: Having determined the allowable cost of capital the regulatory commission "multiplies the allowable cost of capital by the rate base to determine the allowable after-tax revenues. The commission adds tax expenses (ignoring the tax shield due to interest payments), estimated production costs [presumably including depreciation], and other expenses to these after-tax revenues. The resulting figure is analogous to the net sales of a manufacturer. The commission divides this figure by the forecasted demanded quantity, and the result is the allowable price".

12. This can be checked by combining these equations to show that equation 2 still holds.

13. See Schmalensee (1989) and Rogerson (1992).

14. This is no longer true in a context of uncertainty as discussed later.

15. Other forms of depreciation are described in Hardin, Ergas and Small (1999).

16. See, for example, Myers, Kolbe and Tye (1985).

17. As one example of the use of straight-line depreciation applied to the optimised replacement cost, MED (2000) page 19 states that when valuing the assets of electricity line businesses in New Zealand, the following formula is used: DV = UDV x RL/TL where UDV is undepreciated optimised replacement cost, RL is remaining life of the asset and TL is total life.

18. For a detailed description of one such model see NERA (1999).

19. CRNEC writes: "Some optimisation models are based on a 'scorched node' approach, in which the location of key elements of the existing infrastructure are taken as given. Typically this includes the nodes, such as network switches and points of interconnection. Since these locations are fixed, however, the communication cables that join them are also likely to remain in the same place after optimisation. Because less optimisation is undertaken, a scorched node network will generally have a higher cost than a scorched earth network that provides the same level of service. The resulting prices will therefore also be higher. Since there is not a complete mapping between assets used and services provided, it is possible that the relative prices of two services may change over time, but comparing a scorched node and a scorched earth network on any particular date cannot result in overall network valuations for the former that exceed those of the latter." CRNEC (2001), page 13.

20. The implementation of a modified scorched node approach is one of the "principles of implementation and best practice" of the Independent Regulators Group (2000), page 3.

21. Cited by the Productivity Commission (2001), page D.8.

22. Cited by the Productivity Commission (2001), page D.8.

23. An interesting question is whether or not there is an economic justification for avoiding rate shock? Is the desire to avoid rate shock a consequence of the desire to avoid adjustment costs, or is it due to a concern with inter-temporal equity (*i.e.*, is it considered "unfair" that consumers of the same good at different points in time pay different amounts?).

24. If the regulator does not have knowledge of the path of the replacement cost over the entire life of the firm it cannot ensure a continuous path of earnings.

25. Crew and Kleindorfer (1992) show that when the price at which new entrants will enter the market is declining over time, straight-line depreciation may require a level of earnings in later periods which induces inefficient asset duplication.

26. In the case where the regulated asset will be replaced at the end of its life, the relevant expression is only a little more complicated (see the appendix).

27. For more on the Ramsey approach to determining an optimal depreciation path, see Burness and Patrick (1992).

28. β is equal to the covariance of the firm's returns with the market portfolio divided by the variance in the market portfolio.

29. For example, see ACCC (2001) or MED (2000), page 75.

30. Various studies attempt to measure how beta varies over time, see Buckland and Fraser (1999) (for the regulated water industry) and Fraser *et al.* (2000).

31. This last example illustrates that a simple change in the way that retail prices are structured, shifting the risk to which the regulated firm is exposed, can lower its cost of capital and therefore the costs that the end-users must pay.

32. Ergas and Kuypers (2001), page 10.

33. Strictly speaking, marginal cost need only be constant over the relevant range over which output varies – if marginal cost is constant over part of the range and then increases rapidly so that no further output is ever produced (as is the case when the network reaches congestion) the cost function is still constant over the range in which output varies (although in this case the variation in output is truncated).

34. See Ehrhardt (1994), page 170.

35. Under the three-moment model, the required return on equity is given by $r_f + \beta b_1 + \gamma b_2$ where r_f is the "risk free" rate of return; β has the same interpretation as the traditional CAPM beta; γ is a measure of the risk attributable to skewness; and b_1 and b_2 are the risk premiums for the corresponding measures of risk.

36. Conine and Tamarkin (1985), cited in Ehrhardt (1994).

37. See Ehrhardt (1994), page 171.

38. "Carleton (1978) argues that the impact of the regulatory process alters the expected return distribution to the extent that application of the CAPM is no longer valid. In fact, he suggests that no asset pricing model is valid unless that model explicitly takes into account the regulatory process. Other authors, for various reasons, also argue that the CAPM should not be used to estimate the cost of equity for a utility.

On the other hand, many other authors argue that the CAPM can be used. Unfortunately, there is no definitive answer." Ehrhardt (1994), page 169.

39. At least one regulatory authority has argued that it is not necessary to either increase the size of the asset base or increase the allowed rate of return to account for technological obsolescence, provided the investment is depreciated using the "tilted annuity" approach. This is not correct. Any undepreciated capital carried over into later periods runs the risk that it will not be able to be recovered because demand has disappeared. In other words, the tilted annuity approach can only work if the annuity is tilted to a very extreme extent – so that all the depreciation is in the first period.

40. Note that in this case the market value of the firm will always be lower than the "book value" of the firm (given by the level of asset base).

41. Hausman (2001), paragraph 44. As noted later, considerable demand uncertainty is more likely to be an issue for new services (such as "3G" services) than for conventional voice-over-fixed-network services.

42. In some cases it may be possible to subsidise investment in risky assets through other mechanisms (such as subsidies for R&D). To the extent that investments are subsidies in other ways there is, of course, less need to reflect investment risks in a higher cost of capital or asset base.

43. Unfortunately, as has been said, "it is twice as hard to crush a half-truth as a whole lie".

44. According to Burness and Patrick (1992), a depreciation schedule of the kind illustrated here would, in fact, not be allowed under some accounting rules. According to these authors, Financial Accounting Standards Board rules are that recovery not be "backloaded", implying that recovery in any period cannot exceed that in the preceding period.

45. This approach is used, for example by the ACCC in its "Post-tax Revenue Model". Rogerson (1992) calls this "Real Straight Line" depreciation.

Glossary of Terms

ACCC	"Australian Competition and Consumer Commission" – Australia's competition authority and telecommunications regulator.
ADSL	See DSL.
bill-and-keep	An pricing scheme for the two-way interconnection of two networks under which the reciprocal call termination charge is zero – that is, each network agrees to terminate calls from the other network at no charge.
busy hour	The 60 minute period during which a telecommunications network experiences its peak traffic.
CBD	"Central Business District" – The central, built-up region of larger towns and cities which usually has the highest geographic density of telecommunications services.
CLEC	"Competing Local Exchange Carrier" – A licensed local operator in the US and Canada which is not the incumbent local operator, or ILEC. See ILEC.
CMRS	Commercial Mobile Radio Service – A US FCC designation for any carrier or licensee whose wireless network is connected to the public switched telephone network and/or is operated for profit.
Complementary	For the purposes of this paper two or more services are complementary when both or all are required to obtain the final combined or bundled service that the end-user wants. For example, car bodies, car engines and car tires are complementary – all must be combined to obtain the final service that consumers want (in this case car transportation services).
Competitive	A service is competitive if the market for the provision of that service can sustain a high level of competition – that is, many simultaneous competing operators. Compare with non-competitive.
CPP	"Calling Party Pays" – The convention under which the party originating a call pays the entire end-to-end cost of the communication. In particular, in the mobile sector, this implies

	that the calling party pays the "airtime charge" for termination on the mobile handset. Compare with RPP.
CPI	"Consumer Price Index". See RPI.
CRTC	"Canadian Radio-Television and Telecommunications Commission". See *www.crtc.gc.ca*.
DORC	"Depreciated Optimised Replacement Cost" – An approach to allocating the capital costs of a project under which the regulatory asset base is periodically "re-valued" to be equal to the price of building or buying a modern equivalent asset, depreciated to reflect the shorter remaining life of the existing assets.
DSL	"Digital Subscriber Line" – One of a range of technologies which allows high bandwidth services to be carried over a traditional copper twisted pair. The bandwidth achievable usually depends on the length of the copper pair. The most common form is ADSL or "Asymmetric Digital Subscriber Line" under which the bandwidth available in each direction is different. These technologies are collectively known as xDSL.
erlang	Standard unit of measurement of telecommunications traffic capacity and usage demand. Used in traffic engineering for throughput and capacity planning. One erlang equals one standard voice circuit in use full time. See the telecommunications dictionary –*www.ctechsolutions.com/dictionary.htm*.
FCC	"Federal Communications Commission". See *www.fcc.gov.*
FRIACO	"Flat Rate Internet Access Call Origination" – Access to call origination services for calls to internet service providers which is charged at an unmetered or flat rate.
ILEC	"Incumbent Local Exchange Carrier" – A licensed local operator in the US and Canada which historically operated all or almost all the local loops in a geographic region. Compare with CLEC.
ISDN	"Integrated Services Digital Network" – A digital local loop connection to the nearest exchange. The bandwidth is not as high as DSL or cable local loops.
ISP	"Internet Service Provider" – A company which provides end-users with a data connection allowing access to the internet and the associated services (World Wide Web, Email, Chat rooms, Instant Messaging, Internet Telephony and so on).
IXC	"Inter-eXchange Carrier" – An operator in the US and Canada which is licensed to provide long-distance service between two local areas.

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

KPN	Incumbent telecommunications operator in the Netherlands. See *www.kpn.com*.
LEC	"Local Exchange Carrier" – See ILEC and CLEC.
Local loop	The connection between the end-user's premises and the nearest local exchange, multiplexor or concentrator. The local loop is usually in the form of a pair of copper wires, but could also be a fibre-optic cable or a co-axial cable.
LLU	"Local loop unbundling" – The provision of access to both physical ends of the local loop (in some cases access to the local loop is shared between two companies, as when one company provides data services and another voice services over the same local loop).
Non-competitive	A service is non-competitive if the market for provision of that service can only sustain a single firm. Usually this is due to the presence of economies of scale – *i.e.*, the service is a natural monopoly. It may also be due to regulation. Compare with competitive.
NRA	"National Regulatory Authority" – in the EC directives, the regulatory institution responsible for issuing licences, controlling prices, resolving disputes, etc.
One-way call origination	The local loop service connected to the originating party provided by network A to network B where network B does not also provide reciprocal services to network A. For example, the origination service provided by a LEC to originate long-distance calls carried by an IXC in the US, or call origination provided by a ILEC to an ISP.
One-way call termination	The local loop service connected to the receiving party provided by network A to network B where network B does not also provide reciprocal services to network A. For example, the termination service provided by a LEC to terminate long-distance carried by an IXC in the US, or possibly the termination service provided by a paging company for a call from an ILEC to a paging company.
POI	"Point of Interconnection" – the geographical location where two networks interconnect and exchange traffic.
price cap	A system for regulating the prices of a bundle of services of a regulated firm under which the individual price for each service is not controlled but there is a ceiling on the weighted average of all the prices in the bundle.
Retail-Minus	An approach to setting access prices under which the access prices are explicitly set on the basis of the end-user or retail prices

	of the corresponding final services. The discount off retail prices is usually set as a fixed percentage of the retail price.
RIO	"Reference Interconnection Offer" – Under the existing EC directives, a firm which is designated as having SMP must regularly produce a document the terms and conditions at which it will provide access to specified services. This document must be approved by the regulator.
RPP	"Receiving Party Pays" – The convention under which the party receiving a communication pays all or most of the end-to-end cost of the communication. In particular, in the mobile sector, this is used to refer to the case where the receiving party pays the "airtime charge" for termination on the mobile handset. In this case the originating or calling party may still pay for a local call. Compare with CPP.
RPI	"Retail Price Index" – A measure of the rate of inflation in the goods purchased by consumers. Also known asCPI.
RPI-X	An approach to regulating prices under which the regulated firm is allowed to adjust its own prices subject to the weighted average of prices not exceeding a cap. In the RPI-X price cap system this cap is allowed to increase at the rate of inflation (RPI) less some "X factor" to account for productivity gains or to reduce the regulated firm's rents.
SMP	"Significant Market Power" – Under the current EC Directives, an operator designated as having SMP is subject to specific obligations such as the requirement to produce an RIO and the obligation to have cost-oriented tariffs (except for mobile operators). An operator is presumed to have SMP if it has more than 25% of a telecommunications market in the geographic area in which it is allowed to operate.
TELRIC	"Total element long run incremental cost" – the long-run incremental cost of providing a given network element.
TSLRIC	"Total service long run incremental cost" – the long-run incremental cost of providing a given end-user service (which may use several different network elements).
Two-way call origination	The call origination service used to originate a call on network A which is terminated on network B *and* there is a reciprocal relationship between the networks. The payment for this service is usually not set directly but is, in effect, the difference between the end-user price (the payment for the whole service) and the two-way termination charge (the payment for terminating).

ACCESS PRICING IN TELECOMMUNICATIONS – ISBN 92-64-10592-1 – © OECD 2004

Two-way call termination	The call termination service used to terminate on network B a call which was originated on network A *and* there is a reciprocal relationship between the networks (each needs the other to terminate calls). The clearest example is the interconnection of two local fixed networks.
ULL	"Unbundled Local Loop" – The provision of access to both ends of the copper local loop on a permanent basis, allowing the installation of equipment for upgrading the local loop to provide DSL services or the lease of any such equipment which is already installed.
xDSL	See DSL.

References

ACCC (1999), *Statement of Principles for the Regulation of Transmission Revenues*, 27 May 1999, *www.accc.gov.au/electric/regulation/Statement_PR/pdf/statement.PDF*.

ACCC (2000), "Pricing Methodology for the GSM Termination Service: Draft Report", ACCC, December 2000, *www.accc.gov.au/telco/fs-telecom.htm*.

ACCC (2001), *Post-tax revenue handbook*, October 2001, *www.accc.gov.au//gas/br_reg_iss/post_tax_revenue_handbook.pdf*.

Anderson, Robert D., Abraham Hollander, Joseph Monteiro, W. T. Stanbury (1998), "Competition Policy and Regulatory Reform in Canada, 1986-1997", *Review of Industrial Organization*, 13, 1998, 1777-204

Armstrong, Mark (2001), "The Theory of Access Pricing and Interconnection", Forthcoming, *Handbook of Telecommunications Economics*.

Baumol, W., (1971), "Optimal depreciation policy: pricing the products of durable assets", Bell *Journal of Economics and Management Science*, 2, 638-656, 1971.

Bonbright, James C., Albert L. Danielsen, David R. Kamerschen, (1988), "Principles of Public Utility Rates", Public Utility Reports, Inc., Arlington, Virginia, 1988.

Bower, D., R. Bower and D. Logue, (1984), "Arbitrage Pricing Theory and Utility Stock Returns", Journal of Finance, September 1984, 1041-1054.

Buckland, Roger and Patricia Fraser, (2001), "Political and Regulatory Risk in Water Utilities: Beta Sensitivity in the UK", *Journal of Regulatory Economics*, 19(1), January 2001, 5-25 *www.abdn.ac.uk/accountancy/web_pgs/public/dept/Wps/99rb2.doc*.

Burness, H. Stuart and Robert H. Patrick, (1992), "Optimal Depreciation, Payments to Capital and Natural Monopoly Regulation", *Journal of Regulatory Economics*, 4, 35-50, 1992.

Carleton, W. T. (1978), "A Highly Personal Comment on the Use of the CAPM in Public Utility Rate Cases", *Financial Management*, 7(3), 57-59, Autumn 1978.

Carter, Michael and Julian Wright (1999), "Interconnection in Network Industries", *Review of Industrial Organisation*, 14, 1-25.

Carter, Michael and Julian Wright (2001), "Asymmetric Network Interconnection", mimeo, 2001, *http://yoda.eco.auckland.ac.nz/iep/papers/carter_wright_2001.pdf*.

Commerce Commission (2002), *Interconnection Pricing Methodology*, prepared for the New Zealand Commerce Commission by Frontier Economics, 5 April 2002, *www.comcom.govt.nz/telecommunications/pdf/Interconnection%20Pricing%20Methodology.pdf*.

Conine, T.E., Jr. and M. Tamarkin. (1985), "Implications of Skewness in Returns for Utilities' Cost of Equity Capital", *Financial Management*, 14(4), 66-71, 1985.

Crew, Michael A. and Paul R. Kleindorfer, (1992), "Economic Depreciation and the Regulated Firm and Competition and Technical Change", *Journal of Regulatory Economics*, 4, 51-61, 1992.

EC (2001), "Seventh Report on the Implementation of the Telecommunications Regulatory Package", COM(2001) 706, November 2001, *http://europa.eu.int/information_society/topics/telecoms/implementation/annual_report/7report/index_en.htm*.

EC (2002), "Eighth Report on the Implementation of the Telecommunications Regulatory Package", COM (2002) 695, December 2002, *www.europa.eu.int/information_society/topics/telecoms/implementation/annual_report/8threport/index_en.htm*.

Ergas, Henry, Jeremy Hornby, Iain Little and John Small, (2001), "Regulatory Risk", presented at ACCC Regulation and Investment Conference March 2001, *www.necg.com.au/paperpub/papers-ergas-regrisk-mar01.pdf*.

Ergas, Henry and Tim Kuypers, (2001), "Productivity Commission draft report: telecommunications competition regulation: NECG submission", 6 August 2001, *www.necg.com.au/pappub/papers-necg-prod-com-aug01.pdf*.

Ehrhardt, Michael C. (1994), The Search for Value: Measuring the Company's Cost of Capital, Harvard Business School Press, 1994.

FCC (1996), "First Report and Order on the Local Competition Provisions of the Telecommunications Act", FCC 96-325, *http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-96-325A1.pdf*.

FCC (2001a), "NPRM: In the Matter of: Developing a Unified Intercarrier Compensation Regime", 27 April 2001, FCC 01-132, *www.fcc.gov/Bureaus/Common_Carrier/Notices/2001/fcc01132.doc*.

FCC (2001b), "Order on Remand and Report and Order: In the Matter of: Intercarrier Compensation for ISP-Bound Traffic", 27 April 2001, FCC 01-131, *www.fcc.gov/Bureaus/Common_Carrier/Orders/2001/fcc01131.pdf*.

Fisher, Franklin M. and John J. McGowan, (1983), "On the Misuse of Accounting Rates of Return to Infer Monopoly Profits", *American Economic Review*, 73(1), 82-97, March 1983.

Fraser, Patricia, Foort Hamelink, Martin Hoesli and Bryan MacGregor, (2000), "Time-varying Betas and the Cross-Sectional Risk-Return Relation: Evidence From the UK", *Aberdeen Papers in Accountancy, Finance and Management*, Working Paper 00-9, 2000.

Hardin, Alexis, Henry Ergas and John Small (1999), Economic Depreciation in Telecommunications Cost Models, July 1999, *www.necg.com.au/pappub/papers-hardin-ergas-small-IECecondepreciation-jul99.pdf*.

Hausman, Jerry, (1999), "The Effect of Sunk Costs in Telecommunications Regulation" in *Real Options: The New Investment Theory and its Implications for Telecommunications*, Kluwer Academic Publishers, 1999.

Hausman, Jerry, (2000), Declaration of Professor Jerry Hausman to the ACCC on Pricing of Mobile Services.

Hotelling, H., (1925), "A General Mathematical Theory of Depreciation", *Journal of the American Statistical Association*, 20, 340-353.

Independent Regulators Group, (2000), "Principles of implementation and best practice regarding FL-LRIC cost modelling", *www.agcom.it/eng/documents/PIBs_on_LRIC.pdf*.

Laffont, Jean-Jacques and Jean Tirole (2000), *Competition in Telecommunications*, MIT Press, 2000.

Laffont, Jean-Jacques, Patrick Rey and Jean Tirole, (1998), "Network Competition: Price Discrimination", *Rand Journal of Economics*, 29(1), Spring 1998, 1-37.

LECG (1999), "Historic Cost and Replacement Cost: Efficiency Implications of their Use in Price Setting", by Tony van Zijl and Timothy Irwin, 14 August 2001, *www.comcom.govt.nz/price/Airfield/draftsubs/WIAL_rpt4.pdf*.

Lewis, Tracy R. and David E.M. Sappington (1988a), "Regulating a Monopolist with Unknown Demand", *American Economic Review*, 78, 1988, 986-998.

Lewis, Tracy R. and David E.M. Sappington (1988b), "Regulating a Monopolist with Unknown Demand and Cost Functions", *Rand Journal of Economics*, 19(3), Autumn 1988, 438-457.

MED (2000), *Handbook of Optimised Deprival Valuation of System Fixed Assets of Electricity Line Businesses*, New Zealand Ministry of Economic Development, 4th edition, October 2000, *www.med.govt.nz/ers/inf_disc/odv04/*.

Mercury (1993), "Interconnection and Accounting Separation", submission by Mercury Communications to Oftel, July 1993.

Myers, Stewart C., A. Lawrence Kolbe and William B. Tye, (1985), "Inflation and Rate of Return Regulation", *Research in Transportation Economics*, 2, 83-119, 1985.

Newbery, David (1997), "Determining the regulatory asset base for utility price regulation", *Utilities Policy*, 6(1), 1-8, 1997.

NERA (1999), *Estimating the Long Run Incremental Cost of PSTN Access*, January 1999, *www.accc.gov.au/telco/nera.zip*.

OECD (1999), *Regulatory Reform in Mexico*, 21 December 1999, *www.oecd.org/competition.*

OECD (2000), *Cellular Mobile Pricing Structures and Trends*, DSTI/ICCP/TISP(99)11/FINAL, 16 May 2000, *www.oecd.org/oecd/pages/documentredirection?paramID=20419&language=EN&col=OECDDCoreLive*.

OECD (2001a), *Interconnection and Local Competition*, DSTI/ICCP/TISP(2000)3/FINAL, 7 February 2001, *www.oecd.org/oecd/pages/documentredirection?paramID=3019&language=EN&col=OECDDCoreLive.*

OECD (2001b), *Regulatory Reform in Telecommunications: Canada*, Chapter 7 in the *Regulatory Reform Review of Canada*, Forthcoming.

OECD (2001c), *Communications Outlook* 2001, Paris, 2001, *http://oecdpublications.gfi-nb.com/cgi-bin/OECDBookShop.storefront/220658257/Product/View/932001021P1.*

OECD (2001d), *Restructuring Public Utilities for Competition*, Paris, August 2001, *www.oecd.org/competition.*

OECD (2002), *Competition and Regulation Issues in Telecommunications*, No. 38 in the Series Best Practice Roundtables on Competition Policy, 1 February 2002, *www.oecd.org/competition.*

Oftel (1999), "Access to Bandwidth: Delivering Competition for the Information Age", A statement issued by the Director General of Telecommunications, November 1999, *www.oftel.gov.uk/publications/1999/competition/a2b1199.htm*.

Oftel (2000), "Determination of a dispute between BT and MCI WorldCom concerning the provision of a Flat Rate Internet Access Call Origination product (FRIACO)", 26 May 2000, *www.oftel.gov.uk/publications/internet/fria0500.htm*.

Oftel (2001a), "Oftel's 2000/01 effective competition review of dial-up Internet access", 30 July 2001, *www.oftel.gov.uk/publications/internet/imr0701.pdf*.

Oftel (2001b), "Review of the charge control on calls to mobiles", 26 September 2001, *www.oftel.gov.uk/publications/mobile/ctm0901.pdf*.

Productivity Commission (2001), *Telecommunications Competition Regulation*, Report No. 16, AusInfo, Canberra, September 2001.

Rogerson, William P. (1992), "Optimal Depreciation Schedules for Regulated Utilities", *Journal of Regulatory Economics*, 4, 5-33, 1992.

Salinger, Michael, (1998), "Regulating Prices to Equal Forward-Looking Costs: Cost-based Prices or Price-based Costs?", *Journal of Regulatory Economics*, 14, 149-163, 1998.

Schmalensee, Richard (1989), "An Expository Note on Depreciation and Profitability Under Rate-of-Return Regulation", *Journal of Regulatory Economics*, 1, 293-298, 1989.

Touche Ross (1994), "Capacity Based Charging for Interconnect: Final Report", prepared for Oftel, November 1994.

Valletti, Tommaso M. and Antonio Estache (1998), "The Theory of Access Pricing: An Overview for Infrastructure Regulators", paper prepared for the World Bank Institute, March 1998, *www.worldbank.org/html/dec/Publications/Workpapers/wps2000series/wps2097/wps2097.pdf*.

OECD PUBLICATIONS, 2, rue André-Pascal, 75775 PARIS CEDEX 16
PRINTED IN FRANCE
(24 2004 01 1 P) ISBN 92-64-10592-1 – No. 53333 2004