

Mobile Termination: How to Regulate or Perhaps Not to Regulate at All?

ALF VANAGS

TeliaSonera Institute Discussion Paper No 8

Mobile Termination: How to Regulate or Perhaps Not to Regulate at All?

By Alf Vanags¹

September, 2010

¹ Alf Vanags (alf@biceps.org), Director, Baltic International Centre for Economic Policy Studies (BICEPS), www.biceps.org.

Acknowledgements

I am indebted to Anders Paalzow for extensive and helpful discussions on the theory of termination and its welfare implications as well as for references to the literature and considerable editorial input. I also gratefully acknowledge valuable research assistance performed by Krisjanis Krustins.

ISBN 978-9984-842-23-3

© Stockholm School of Economics in Riga and Baltic International
Centre for Economic Policy Studies (BICEPS), 2010

Table of Contents

- Executive Summary 6
- 1. Introduction 7
- 2. The current situation and the way forward 10
- 3. A short primer on the economic logic of regulation 14
 - 3.1. The role of prices 14
 - 3.2. Market power 15
 - 3.3. Private/social cost divergence: externalities 18
 - 3.4. Competition 19
- 4. Mobile call termination and consumer welfare 20
 - 4.1. The channels of impact of MTRs 22
 - 4.2. Relationship between MTR and retail prices 22
 - 4.3. The revenue channel of impact 24
- 5. Estimates of the consumer benefit impact of changes in MTRs 27
- 6. Cost based charge regulation vs. mandated bill and keep 30
- 7. Conclusions 32
- References 34

Foreword

This is the eighth of the TeliaSonera Institute Discussion Papers series. The Institute, which is located at the Stockholm School of Economics in Riga is generously supported by TeliaSonera and aims to promote applied economic research in the fields of entrepreneurship and telecommunication – the latter with a focus on regulatory issues.

The current discussion paper by Alf Vanags addresses the theoretical and empirical issues of how or whether to regulate the charges made by mobile operators for terminating calls in their network. The current as well as the previous seven discussion papers can be downloaded from the SSE Riga website, www.sseriga.edu.lv. Hard copies can be ordered from office@sseriga.edu.lv.

Anders Paalzow
Rector, SSE Riga

Alf Vanags
Director, BICEPS

Executive Summary

A combination of factors, such as the EU regulatory timetable and the growing convergence to a multi-service Next Generation Network (NGN) IP-network have led to much discussion and examination of alternative mobile termination or interconnection regimes. This paper attempts to cut through a literature that is either quite abstract or reflects the special interests of one group or another. The aim is to provide an independent assessment of the economics of the literature on alternative regimes and to make a judgment as to where the balance of benefit lies from the point of view of society as a whole.

A typology of regulatory regimes is offered in terms combinations of price control and rules. The prevailing regime can be categorized as price control with a minimum of rules. By contrast a mandatory Bill and Keep (BAK) regime has price control (termination rates set at zero) and rules (i.e. symmetry is imposed), whereas a negotiated BAK has symmetry rules but no price control. No rules and no price control is both logically possible, has been adopted in Hong Kong and is not inconceivable as a medium term way forward.

The current situation is described together with some discussion of how we got there and of the feasible directions forward. An extensive discussion of the underlying theory of regulations is linked to the problems that arise in applying regulation in the specific situation of mobile termination. The key analytical proposition here which helps understanding of the impact of regulation of interconnection charges is that the charge enters the decision-making of the mobile network operator in two ways: as a cost that needs to be covered and as a revenue that is sought. This dual character of mobile termination charges, especially when they are set in excess of marginal cost, leads to a number of potential distortions e.g. on-net/off-net differentials that do not reflect costs or excessive and cross-subsidised competition for subscribers.

As an alternative to the prevailing European cost based system of regulating a regime of mandatory BAK has been proposed. The advantages are seen as lower prices (revenue per minute), higher usage (minutes of use), less tariff mediated distortion, and perhaps not least – much less regulation and hence regulatory cost and regulatory failure.

It should be said that mandatory BAK has nowhere been imposed – where it prevails it has emerged through commercial negotiation given the requirement of symmetry. So caution is needed. Nevertheless, the gains from ever tighter cost based regulation are very likely diminishing and a regime switch looks attractive.

A reasonable transition period should allow time to address and resolve certain technical issues like defining the BAK domain and for operators make what will be considerable changes to prices and practices. So, if a decision to switch to BAK were made by end 2012, then the start of 2016 might allow sufficient time for the needed adaptation.

1. Introduction

Over the years there has been persistent concern about the level of wholesale mobile phone call termination charges in countries which use the Calling Party Network Pays (CPNP) convention (see e.g. Littlechild 2006). While there is a common belief that retail call origination has become more and more competitive, both regulators and academics have focussed attention on wholesale call termination, in particular call termination off-net, i.e. outside the network of the originating mobile phone call, as an area of potential market failure². The termination charges that mobile phone operators levy on each other and on fixed network operators are perceived by many (including regulators and policy makers) to be too high if left to the discretion of the operators³, resulting in both excessive profits and inefficient resource allocation. This is because under CPNP the terminating operator can exploit its monopoly in termination of calls in its network and is therefore effectively unconstrained in setting the termination charge.

Once this phenomenon was recognized, the response by European regulators has been to implement a variety of cost based termination rate regulation models. These cost based Mobile Call Termination (MCT) models are now under scrutiny throughout the European Union. This reflects the perception among regulators and policy makers that while the current models have “worked” in the past, accumulating evidence on termination rates and practices suggest that the observed diversity of termination models may not be justified in terms of economic and regulatory efficiency. Rethinking has also been prompted by the evidence on the emerging convergence of fixed, mobile and Next Generation Networks (NGN) which may make the existing regulatory solutions redundant.

The result is that a move on the termination regimes in Europe is imminent and very likely implemented by the end of 2011. This is the time by which according to the EC Recommendation (2009b), termination rates set by European national regulators (NRAs) should be “applied at a cost-efficient, symmetric level”. In the UK Ofcom is considering six alternatives, including Bill and Keep (BAK), as possible new regulatory models for 2012 and the European Regulators Group (ERG) has more or less recommended a switch to a BAK for the long term.

In terms of economic analysis, changing regulatory regime is not a trivial issue. Indeed this is what this paper is about. Furthermore, as the following examples show, there are substantial sums of money involved – in terms of revenues and costs associated with termination as well as in terms of costs associated with regulation and the litigation that usually follows regulation.

- According to Ofcom (2009), mobile termination generates 14% of the revenues of UK mobile operators.
- According to WIK-Consult (2008), it is estimated that fixed operators paid mobile network operators around EUR 10 billion for termination during the period 1998–2006.

² On-net termination is not regarded as an issue for concern because it is internal to the network.

³ However, as Armstrong and Wright (2008) point out, while there is a presumption that the competitive bottleneck arising from a network's monopoly of termination in its own network gives rise to an incentive to raise termination charges above marginal cost, when mobile networks compete there may be an incentive for them to jointly set mutual termination rates below marginal cost so as to soften competition in the retail market.

- According to the European Commission (2009a), the approach suggested in the EC Recommendation would result in a EUR 4 billion transfer from mobile operators to consumers and fixed operators (about equally) over the period 2009–2012.
- The cost of regulating mobile termination charges amounts, according to Littlechild (2006, p. 271) to approximately USD 300 million per year in Western Europe alone. In addition there are most likely substantial costs associated with litigation.

It is useful to think of termination regimes in terms of the instruments at the disposal of the regulator: one instrument is the wholesale price for mobile call termination; a second is the set of rules applied to commercial termination agreements e.g. mandatory symmetry or reciprocity of rates. Other instruments may also be relevant e.g. connectivity requirements, but here we concentrate on the first two. If we think of the instruments as independently applied then this gives rise to four logically possible types of potential regulatory regimes. These are illustrated in Table 1, together with examples of regimes corresponding to each type.

	Price control	No price control
Rules on termination agreements	Mandatory bill and keep (BAK)	Symmetry imposed on commercially negotiated termination rates Negotiated BAK
No rules	Price cap: may be different for different operators, e.g. LRIC+, LRMC, and capacity based pricing	No regulation Market based agreements

Table 1. Typology of mobile termination regulatory regimes⁴

Thus the current cost based wholesale termination regime applied in most of Europe can be identified as a 'price control' and 'no rules' combination. Although the models differ from country to country, most countries apply a version of price capping and most are also based on the concept of Long Run Incremental Cost (LRIC) although some countries use other cost benchmarks. LRIC is defined as the costs caused by the provision of a defined increment of termination at a given level of volume. The LRIC+ concept applied in many European countries is an average cost based pricing methodology and hence also includes a provision to cover a proportion of common/fixed costs (usually through a mark-up). Long Run Marginal Cost (LRMC), on the other hand, does not take common/fixed costs into account and considers smaller increments in volume. The cost on which the price cap is based can be derived from costs of each individual operator or it can be benchmarked. Furthermore, the regulator can choose a uniform price cap or a discriminatory one. Another regulatory regime that belongs to the 'price control' and 'no rules' set is Capacity Based Charges (CBC), which is used in Canada⁵. With CBC, prices are set equal to the incremental cost of the capacity supplied to the interconnecting party, i.e. a fixed fee for a specified network capacity⁶. Irrespective of whether LRIC or CBC is employed, the regulator has

⁴ It should be mentioned i) that the typology applies to wholesale regimes only – retail prices depend in a possibly complicated way on the wholesale regime. Also the term rules, as used here, applies to rules on the structure of termination agreements. Thus a 'no rules' regime does not, for example, exclude mandatory requirements to maintain connectivity or other regulatory requirements not directly connected with termination agreements.

⁵ For a description see Analysis Mason (2008) pp 11–13.

⁶ Various aspects of LRIC are discussed in Ofcom (2009) and Post och Telestyrelsen (2002), whereas capacity-based pricing is discussed in Kennet and Ralph (2007).

to have an underlying model of the operators' cost structure and feed it with data. This is far from a trivial exercise and hence, as discussed above, costly to administer not only as such, but also in terms of the litigation it tends to generate.

On the other hand, the 'rules' but no price control corresponds to proposals that the level of termination rates should not be directly controlled by regulators but that commercial agreements should be subject only to symmetry i.e. between mobile and fixed operators or mobile and mobile. This is the system that formally applies to mobile termination in the US. Interestingly in many cases this has resulted in commercially negotiated BAK arrangements.

The mandatory bill and keep (BAK) alternative is a combination of price control (in this case zero termination rates) and symmetry between operators.

No regulation is potentially an interesting option that has recently been adopted in Hong Kong (since April 2009) but does not yet appear to be regarded with much favour by European regulators. However, this is one of the cases formally considered by Ofcom (2009) and is of interest because of convergence to Internet Protocol interconnection and which currently is not subject to regulation. Moreover, this seems to work where commercial considerations have resulted in a hierarchical system of peering for same size networks and transit charges for traffic between asymmetric networks.

The current regulatory debate suggests that realistically there are two medium term alternatives for EU termination regimes:

- The current Calling Party Network Pays (CPNP) convention but with a tighter price cap;
- Mandatory Bill and Keep (BAK).

These two alternatives will be the focus of the forthcoming discussion. Already now it should be noted that the two regimes generate different marginal costs to the operator when it comes to on-net and off-net calls. Under the CPNP regime the marginal cost for an off-net call (i.e. a call terminating on another operator's network) is most likely higher than for an on-net call. A change in the price cap will accordingly change the marginal cost. This might in turn result in termination-based price discrimination at the retail level, i.e. the operator sets higher prices for calls terminating off-net. On the other hand, were the prevailing regime BAK, then the whole marginal cost for calls on-net and off-net would be similar (or very close to similar). Accordingly, termination-based price discrimination would be less likely.

The rest of the paper is organized as follows. The next section addresses the institutional framework with an emphasis on the European Union. Section 3 provides the basic economic theory behind regulation and is followed by a section on the interpretation of the basic theory in the context of mobile call termination and consumer welfare. Section 5 reports estimates of the costs and benefits of changes in mobile termination regimes. Section 6 discusses and summarises the relative merits of cost based charge regulation versus bill and keep. The last section has some concluding remarks.

2. The current situation and the way forward

Currently, wholesale mobile termination rates are regulated everywhere in Europe. This is motivated by the fact that the network in which a call is terminated effectively has a natural monopoly in the provision of termination. As discussed in the previous chapter, the current European regime is based on the Calling Party Network Pays (CPNP) principle at the wholesale level. At the European retail level this has translated into the Calling Party Pays (CPP) principle which in turn may create incentives for termination-based price discrimination. Thus, the CPNP regime creates an opportunity for the terminating operators to exploit the monopoly power that arises from the bottleneck monopoly that comes with termination. This is likely to have negative implications for economic efficiency in general and for the economic welfare of the consumers of mobile telephony in particular. Furthermore, as discussed in Littlechild (2006), the bottleneck monopoly and the possibility to exploit it within a CPNP regime has resulted in a widespread concern at the level of termination charges, which in turn has given rise to regulatory measures addressing the price of off-net termination.

The general argument for why mobile termination is a problem under the CPNP convention is well summarized by the European Commission (2009a, p. 4):

“As the termination charge is set by the called network, which is chosen by the called subscriber, the calling party in general does not have the ability to affect or influence the level of those termination charges... as call termination to a geographic number can be supplied only by the network provider to whose network the called party is connected, there is neither a demand- nor a supply-side substitute for call termination on an individual network. Therefore, each network constitutes a separate market and each company is considered to be a monopolist on its own network.

As a consequence, unregulated fixed or mobile operators would have an incentive to set their termination charge at the monopoly level. The CPP convention allows the terminating operator to raise its prices typically without a constraint from either party to the call. The calling party pays a bundled fee, including the termination rate, to its network operator at the retail level and *will therefore not generally see a direct price signal for the wholesale termination service.*⁷ As the receiving party makes no payment for the termination service by convention (CPP), it generally has little or no incentive to constrain the pricing behaviour of its terminating operator. To the extent that the increased wholesale termination charge results in increased retail prices and reduces the number of calls that an end-user receives, they are made worse off. However, this may not be directly perceptible to the end-user such that it cannot necessarily attribute this fall-off in calls to a higher termination rate. Thus, terminating operators have the ability to raise the price of reaching their subscribers substantially above cost.”

⁷ Italics added.

International evidence suggests that under CPNP without a price cap, i.e. when mobile network operators (MNOs) have been unconstrained in setting termination rates these have been set substantially above cost. Table 2 illustrates the termination charge mark-up prior to the introduction of serious price control for selected markets.

Country/year	Markup over cost (%)
UK 1998	22–30
UK 2002	30–40
France 2001	66
Ireland 1999–2003	>43
The Netherlands 2003	100
Australia 2004	>100
New Zealand 2004	75
International (Ovum study) 2004	70
Europe (Ovum study) 2004	>100

Table 2. Regulatory findings on termination charge markup over cost

Source: Littlechild (2006)

Regulation has of course reduced termination charges. For example in the UK the termination charge paid by BT has fallen from in excess of 25 pence per minute in 1995 to about 5 pence per minute in 2009. Although costs have fallen over this period much of the decline in termination rates very probably has been the result of active regulatory pressure.

So far the discussion has mainly focussed on the wholesale market. However, the wholesale regime is likely to have an impact on the retail level as well. Not only in terms of prices, but also on the retail payment model employed by the operators. In Europe the CPNP regime applies at the wholesale level and almost everywhere the Calling Party Pays (CPP) applies at the retail level⁸. Were the regulator on the other hand, choosing a policy that would result in BAK at the wholesale level (like in the U.S. for example), then this most likely (although not necessarily) will result in a Receiving Party Pays (RPP) as well as CPP regime at the retail level.

As a consequence wholesale mobile termination is nearly everywhere regulated by means of a price cap and it is widely regarded that, despite the fact that interconnection rates in the EU have fallen from an average of 12.65 euro cents per minute in October 2005 to 8.55 euro cents per minute in October 2008, current levels of charges remain well in excess of cost. Thus, for example, in the central scenario of the model accompanying the EU Recommendation (see European Commission (2009a)) it is assumed that even in the absence of a new termination model, normal regulatory pressure could result in a reduction of termination rates by 40% over 2009–2012 while implementation of the new termination model suggested in the EC Recommendation would result in an average termination rate reduction of 70% over the same period.

⁸ An exception is international roaming where normally the receiving party also pays what is usually a substantial charge.

The very large differential between fixed and mobile termination rates in Europe provides further evidence suggesting that despite falling in recent years in Europe mobile termination rates may still be excessively high. Thus, according to International Telecommunications Union data for 2006–2007⁹ mobile termination rates in the Europe Mediterranean Basin were more than 10 times higher than fixed termination rates as against a global average ratio of 3.77 and a North America ratio of 2.16, while in the Asia Pacific region the ratio was lower still at 1.42. It does not seem plausible that such differences could be the result of differences in the underlying costs of fixed and mobile termination.

Not only are termination rates in Europe regarded as high they are also quite dispersed, ranging from 2.10 euro cents in Cyprus to 15.92 euro cents in Bulgaria. The full range of rates is shown in the Figure 1.

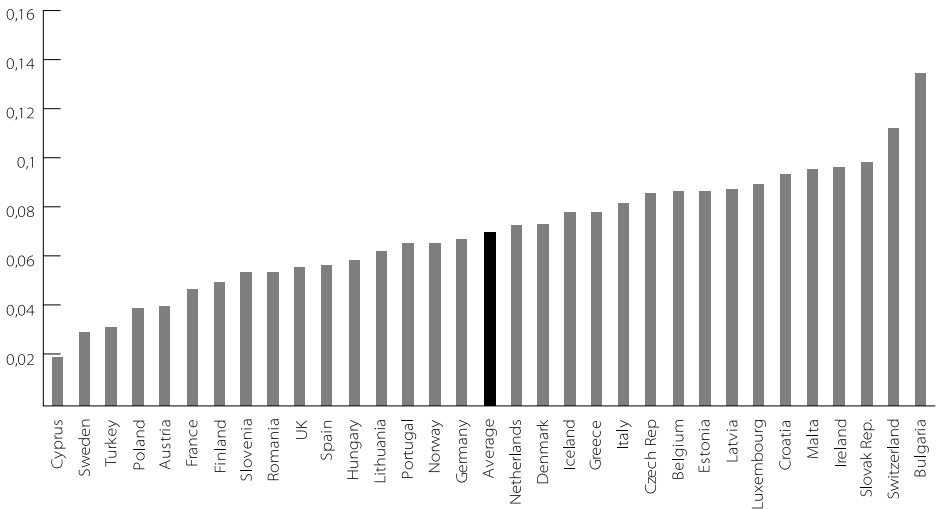


Figure 1. Dispersion of mobile termination rates in Europe July 2009 in euros

Source: European Regulators Group (2009)

The rather wide dispersion in rates is partly a consequence of different actual cost levels across countries but is equally or perhaps more so a consequence of different cost methodologies applied by regulators in different countries.

The perception of excessively high rates combined with the fragmented European approach to cost based price regulation has led the Commission to develop its Recommendation (European Commission 2009b) which aims to achieve two goals: one is to reduce the inconsistencies in the methodologies and practices applied in the regulation of termination rates across the EU member states, thereby reducing distortions in cross border competition and investment; and the second

⁹ See: <http://www.itu.int/osg/spu/presentations/2007/kelly-pricing-billing-interconnection-for-NGN-1-march-07.pdf>.

is to introduce a common cost methodology that will further reduce average levels of mobile termination rates.

The methodology proposed in the Recommendation is to move from the Long Run Incremental Cost + (LRIC+) approach which is the method developed in the UK and currently applied by perhaps two thirds of EU regulators to what is variously labeled as LRIC (i.e. without the +) or Long Run Marginal Cost (LRMC).

The key difference between the prevailing methodology and the one proposed in the Recommendation lies in the (+) which signifies that LRIC+ allows for an element of fixed and common cost recovery. By contrast under LRIC or LRMC there is no allowance for recovery of fixed and common costs.

It is this latter deviation of LRMC from LRIC+ that is responsible for the European Commission (2009a) estimate that adopting the Recommendation would imply “over the period 2009–2012 a potential reduction in cash flows/profits for the mobile industry of €4 billion compared to the baseline scenario. At the same time the fixed sector would gain approx. €2 billion and consumer surplus would also increase by approx. €2 billion”. This claim will be discussed in Section 5 below.

Apart from Commission activity in this area the UK regulator Ofcom has also started a consultation process as has the European Regulators Group, ERG, (2009). Ofcom (2009) in its consultation has identified six possible options for a future termination regime. One is the status quo i.e. LRIC+, the other is the Commission Recommendation i.e. LRMC. The remaining four are:

- **No regulation of mobile call termination.** In practice this has not been common. The main example of non-regulation is Hong Kong where regulation was phased out in 2009.
- **Capacity Based Charges (CBC).** Under this approach termination charges would be set on the basis of the capacity required for terminating traffic. This would change the *structure* of price control, which would very likely to be in the form of fixed monthly or annual charges for a given amount of capacity, in contrast to the per minute charge currently applied. Capacity based termination charges are currently applied in Canada.
- **Mandated Reciprocity (potentially with fixed rates).** Mobile call termination charges could be set at the same level as fixed call termination (termed ‘reciprocity’). This is broadly similar to the approach adopted in the US and in practice appears to have resulted in commercially agreed BAK arrangements.
- **Mandated “Bill and Keep” (BAK).** No payments for termination – that is, termination charges are set at zero.

So what can be realistically expected for the future? There seems to be a certain degree of academic opinion in favour of Bill and Keep and the ERG (2009) has recently produced a consultation

document which appears to very strongly favour BAK as a long term solution. However, most mobile network operators seem to be strongly in favour of keeping the current system, even if it is tightened and rationalized as proposed in the EC Recommendation (2009b). Thus it seems that the leading contenders for the 'next generation of regulatory regime' are tighter price cap regulation on the basis of LRMC and a regime 'switch' to BAK.

3. A short primer on the economic logic of regulation

The literature on regulation, including telecoms regulation, sometime makes it difficult to see the wood for the trees. It is often a curious mixture of economics, law and various populist concepts in which means and ends are sometimes confused and conflated. Thus for example, sometimes, 'competition' appears to be regarded as an end itself whereas in general it should more properly be seen as a means by which desirable market outcomes might be generated. Or, there is sometimes talk of measures having the effect of 'distorting competition' which at once raises the question: "distorting from what?"

The aim of this section is therefore to provide a simple theoretical framework which provides a structure for interpretation of the specific applied literature on telecoms regulation in general and the regulation of mobile call termination in particular. From a theoretical point of view, the economic rationale for regulation is that the market exhibits some type of market failure. In the market for mobile call termination the market failures stem from at least three different sources:

- Existence of a natural monopoly;
- Presence of externalities;
- Information asymmetries at the retail level.

Each of them will be discussed below. However, before the discussing these issues we have to discuss the role of prices in a market economy.

3.1. The role of prices

The starting point is one of the basic propositions of economic theory which asserts that the prices generated in a 'competitive market' i.e. a market where participants have negligible market power¹⁰ will lead to an allocation of resources that is efficient and where no participant earns abnormal or excess profits. Thus producers in a competitive market respond to prices by supplying and trading a quantity such that:

Price reflects marginal cost: this is the condition for an efficient allocation of resources in that resources are allocated to a particular activity up to the point where at the margin consumers

¹⁰ The term a 'competitive' market is meant to signify that participating firms are 'price takers', which in turn means that the actions of individual firms have a negligible impact on market price. This is to be contrasted with the more every day use of the terms competition or competitive which are used as indicators of fierce rivalry.

are just willing to pay for the cost of supplying the marginal unit. Such a price also sends the right signal for evaluating the benefits of investment in the sector or activity.

At the same time freedom of entry in a competitive market ensures that excessive profits are competed away. In other words:

Price just covers average costs (including normal profit).

These two conditions represents the benchmark according to which actual markets structures may be judged – if price is different from marginal cost then we observe ‘*market failure*’ because resources are misallocated in terms of cost vs. willingness to pay and if price is above average cost we observe ‘*excess profits*’. When there is a presumption that market structure is such that these conditions are likely to be violated in an unconstrained market then ex ante regulation may be appropriate in the form of constraints on market participants (remedies) aimed at correcting the deviation from the desirable competitive outcome¹¹.

In practice failure to meet one or both of these benchmark conditions arises from either the presence of market power or a divergence between perceived private and social costs and benefits (externalities).

3.2. Market power

Most simply or commonly, market failure may arise because of the presence of barriers to entry. These may be either natural or may have been created by the enterprise itself and which allow it to raise price above costs (inclusive of normal profit). The effect of this is illustrated in Figure 2.

In Figure 2 it is assumed that costs do not vary with the level of output of the activity, which implies that marginal and average costs are equal ($MC = AC$ in Figure 2). As discussed below, when there are fixed costs or common costs (as is typical in telecoms) this presents particular additional problems for achieving optimal resource allocation either through market forces or through regulation.

Under the constant cost assumption a competitive market would supply the quantity Q^C which would be sold at a price P^C . This price reflects marginal cost and at the same time does not lead to any excess profits because price is equal the average cost (where average cost is taken to include normal profit). This is the benchmark against which actual market outcomes are to be judged.

¹¹ Thus according to Ofcom (2009): “in the absence of regulation, MNOs would have both the incentive and the ability set excessive prices for MCT” (p. 26).

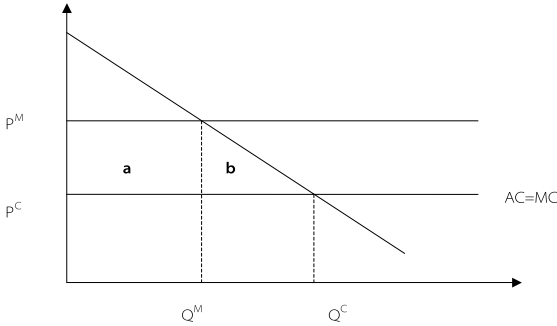


Figure 2. Market failure arising from market power

If an enterprise has market power, e.g. because it is a natural monopoly (like with mobile call termination) or because it can take actions to prevent the entry of competitors, it will not be constrained to the price P^C and in general will have an incentive to raise price above this level and correspondingly to reduce supply¹². Thus price might be increased to P^M and at this price the reduced quantity, Q^M would be supplied. The exact amount by which price might be raised depends on the responsiveness of demand to price and this in turn would depend on the availability of substitutes to the service provided by the given enterprise.

Given that price is set above the competitive level two effects may be identified: one is the presence of *excess profits* and the other is the efficiency or welfare loss associated with the fact that the supply is less than the socially efficient level, Q^C . This efficiency loss is sometimes known as the *deadweight loss*. In Figure 2 the size of excess profits is given by the area **a**, which corresponds to the amount by which price exceeds the competitive level, $P^M - P^C$, multiplied by the quantity supplied (traffic), Q^M . The area **b** represents the loss to consumers (measured in terms of consumers surplus) arising from fact that only Q^M is supplied (and hence consumed) in the market power regime as compared with the efficient level, Q^C .

Note that these two effects are very different, even though both arise from the presence and exercise of market power. The excess profits, area **a**, represent a transfer from one set of agents to another e.g. if Figure 2 represents a retail market, **a** represents a transfer from consumers to the supplier. If on the other hand Figure 2 represents a bottleneck as in off-net wholesale termination then under the calling party network pays (CPNP) regime it corresponds to a transfer from the originating to the terminating network. On the other hand, the area **b** represents 'pure inefficiency' in the sense that it is a measure of what consumers would have been willing to pay over and above cost for consuming the marginal quantity, $Q^C - Q^M$, but which they are prevented from consuming because supply is restricted to Q^M because price is above the competitive level. This is a loss to consumers that does not have a corresponding gain anywhere else, hence the term deadweight loss.

¹² The ability to raise price above the competitive level is precisely what defines a situation of significant market power (SMP) in the electronic communications regulation framework.

The theoretically best remedy for the situation depicted in Figure 2 is to mitigate market power by removing the entry barriers responsible for the market power, thereby relying on competition to drive the price down to the competitive level, P^C . This is, however, not always possible e.g. because the bottleneck is a natural monopoly (like mobile call termination). In this case the normal remedy is to regulate (cap) the price. Hence to regulate MCT under a CPNP regime requires correct identification of the 'competitive price', P^C , an undertaking that is not at all straightforward – even in the simplest case of constant costs as depicted in Figure 2. In the first place, the information required has to be obtained from the operator who has little incentive to reveal costs correctly. Secondly, if marginal costs are not constant then identifying the correct regulation price is much more complicated and depends also on the interaction with demand and the cost schedule, which in turn means that the correct regulation price is not independent of the scale of activity. Hence, the challenge facing the regulator is to estimate the cost to set the price cap equal to P^C in order to minimize the impact of excessive prices, as measured by areas **a** and **b**. This cost (and hence also price) estimate will as discussed in the introductory chapter provide a basis for the price cap under CPNP – applying the concept of Long Run Incremental Cost (LRIC) is for example one way of approaching the problem.

A further complication arises if there are fixed costs. For example, suppose that as compared with the situation depicted in Figure 2, there is a fixed cost associated with supply in addition to a constant variable cost. This situation is depicted in Figure 3.

The presence of fixed or common costs means that average costs are falling with increased supply and are above marginal costs. This in turn implies that pricing at marginal cost (the efficient price) would lead to a loss (of an amount depicted by the shaded area in Figure 3).

The theoretical regulatory solution to this has been either some form of two part tariff, i.e. a fixed charge and variable one set at marginal cost or average cost pricing, depicted by the price P^{AC} in Figure 3. As seen from the figure, average cost pricing eliminates the excess profits generated by market power, but usage remains inefficient with supply and consumption constrained to Q^{AC} , which is below the optimal level Q^C . A deadweight loss remains although it is less than at the monopoly levels of price and supply¹³.

A similar problem arises if the enterprise supplies multiple goods or services and there are joint or common costs. In this case the optimum pricing rule is marginal cost for individual goods or services with a Ramsey¹⁴ mark-up for the recovery of the joint or common costs. The Ramsey rule is to set the mark-up over marginal cost in inverse proportion to the price elasticity of demand for individual goods and services. Needless to say the informational requirement for setting Ramsey prices are very demanding and if based on incorrect information may be inferior to average cost based prices from an efficiency perspective.

¹³ The current European regulation of mobile call termination is a LRIC+ regime which as Ofcom (2009) notes "is an average cost pricing methodology [and] may increasingly lead to retail pricing structures that do not fully reflect the underlying costs of the network". Thus LRIC+ is necessarily a 'second best' regime.

¹⁴ Originally developed by Ramsey (1927).

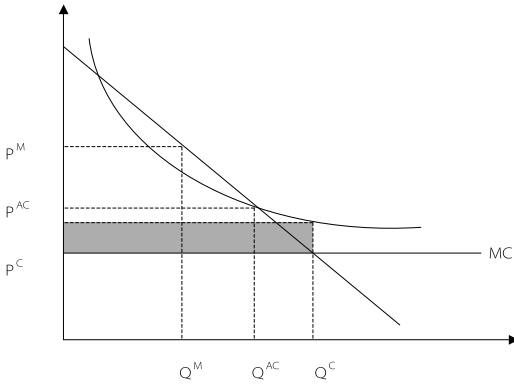


Figure 3. Fixed costs

3.3. Private/social cost divergence: externalities

The optimality of the price equals marginal cost pricing rule presumes that the price reflects all the benefits of the marginal unit of demand and that marginal cost reflects all the costs of the marginal unit of supply. In practice this is often not so i.e. an activity may generate benefits that are not captured in the price received by the supplier or may impose costs that are not reflected in private costs as borne by suppliers. The presence of such *externalities* implies that the market allocation of resources as between activities may be sub-optimal i.e. if an activity generates marginal benefits not captured as revenue by the supplier then the market is likely to undersupply the relevant good or service and, similarly, if an activity imposes costs that are not borne by the supplier then it is likely to be oversupplied. If such externalities are present in a regulated sector then the pricing regime should be designed to reflect them.

A telephone call gives rise to an externality since affects both the caller and receiver – the call receipt externality (see e.g. Albon and York (2006)). Thus if in a Calling Party Pays (CPP) regime there are positive *call externalities*, i.e. the call generates a positive benefit to the receiving party, then in general the calling party will not take this into account when deciding whether to make a call or not. This will result in a less than socially optimal number of calls being made¹⁵. The general regulatory remedy for this type of externality is to ‘internalise’ it and in a regulatory regime the pricing rules should include an externality component in order to achieve this. If the receiving party also pays for the call, i.e. a RPP regime, then this partially or wholly internalizes the call externality. For example, if origination and termination generate equal costs and the caller and receiver value the call equally then an equal charge for the caller and receiver will generate an optimal result. Full internalization in all circumstances is difficult to achieve because it depends on the distribution of benefits as between the caller and the receiver (which is inherently unknown) as well as on the costs generated in both origination and termination.

¹⁵ If calls have a nuisance impact on the receiving party then the opposite conclusions hold.

In network industries, like mobile phone telephony, another form of externality can arise because the entry of a new network participant may affect the benefit of the network to other users. In general, a larger network makes the network more attractive to other users so the *network externality* is positive. In terms of the price equals marginal cost condition, a positive network externality means that the social benefit of an additional participant in the network exceeds the private benefit to the potential joiner. In other words the existing network participants should be will to subsidise the joining costs of new members or enhance in other ways the attractiveness of joining. With the high penetration rates being observed, this externality should be very small¹⁶. Currently new network subscribers are subsidized in many ways, e.g. by handset subsidies when signing up for a subscription. It is unclear whether this is an efficient way of addressing this externality – in particular if it comes at the price of higher termination charges. Moreover, in general a new mobile subscriber generates a network externality to the existing users of all networks, but networks compete to attract subscribers to their own individual network. This phenomenon is exacerbated by the on-net/off-net tariff differential (observed under the CPNP regime), with the result that more resources may be devoted to attracting subscribers, e.g. through subsidised handsets, than is socially optimal and hence with an end result where the penetration rates are ‘too high’ (too many subscribers than socially optimal)¹⁷. It is even further exacerbated by the on-net/off-net tariff differential, or as Laffont, Rey and Tirole (1998) call it termination-based price discrimination. With termination-based price discrimination, it might pay to sign up with an operator having a large customer base (too many subscribers in the largest network(s) than socially optimal). However, in this context there are also local network effects as discussed in Gorrocher and Zirulia (2009). This is due to the observation that individuals typically call a small subset of the overall population of mobile phone users, the operator(s) chosen by the individual’s social network might affect his/her choice of mobile phone operator.

3.4. Competition

The role and importance of competition is widely misunderstood. Firstly, competition is not an end in itself – the importance and relevance of competition arises from its role in generating desirable economic outcomes¹⁸. The textbook definition of competition corresponds to the outcome of a market in which no supplier possesses market power i.e. it is the benchmark against which market outcomes are judged as in Figure 2 above. Where the term is used to signify intense rivalry market outcomes may not always be socially optimal e.g. practices such as predatory pricing or ‘matching prices’ may not always lead to a long run optimal allocation of resources. In mobile telephony, for example, it is not obvious that competition in the form of subsidized handsets leads to a socially optimal allocation of resources. In practice, because entry to mobile telephony is limited by spectrum, the number of players in national markets is quite small. According to European

¹⁶ Where there are capacity constraints on the size of a network it may be that after a certain point a new participant generates congestion costs, a negative externality, on the existing members of the network.

¹⁷ This is discussed in Albon and York (2008) who claim that: “Empirical evidence that handset subsidies are increasingly concentrated on discouraging churn and on encouraging migration to 3G mobile networks – rather than attraction of first-time subscribers – suggests that mobile subscription subsidies are not targeted at internalising various form of externality, and instead are being used as a competitive tool”.

¹⁸ This misunderstanding is even promoted by the regulators themselves. Thus Ofcom (2009) sets out a set of factors that need to be considered in evaluating alternative regimes. These are: ‘economic efficiency’, ‘distributional effects on consumers’, ‘commercial and regulatory consequences’ and ‘competitive impacts’ where the later is defined as “How far do alternatives promote or harm competition among mobile operators? How is competition between mobile and fixed operators affected?” (p. 33).

Commission (2009c) the average market share of the leading two mobile operators has been around 71% from 2004 to 2007, with a fall to 69% in 2008. There is some variation across countries e.g. the UK market has more operators and is therefore likely to be more competitive, while in Cyprus and Slovenia the market shares of the leading operators are respectively about 85% and 60%. In short, in many EU countries the market structure of mobile telephony is effectively a duopoly together with some fringe competitors. This is hardly the textbook 'competitive market'. Moreover, in such markets there may often be scope for the dominant operators to engage in practices that inhibit the ability of smaller operators to compete effectively or that altogether pre-empt entry. Arguably, the widely observed practice of charge differentials for on-net and off-net calls is an example of such a practice.

4. Mobile call termination and consumer welfare

The above framework (or modeling that is based on it) represents the building blocks of how we can identify, estimate and interpret the economic efficiency impact of changes in the mobile termination regime¹⁹. However practical application of the framework faces immense challenges in particular when it comes to analyzing the impact on consumer welfare. The practical problems include:

- In the first place, interconnection or termination is not a final service provided to consumers but an intermediate input that enables the originating operator to terminate calls in another network. The termination charge appears as one factor that determines the retail price of a call and it is the available choices and prices at the retail level that ultimately determine the economic efficiency impact of alternative termination regimes.
- Secondly, the empirical *ex ante* final price of a call as perceived by consumers may not be a well defined concept. Few if any consumers of mobile voice services face a simple linear price per minute of call. Revenue per minute (RPM) is often used as a proxy for retail price but may give very little indication of the structure of incentives facing people when making choices regarding their consumption of mobile services. In particular RPM is an *ex post* average indicator, whereas in practice the marginal cost (to the consumer) of making a given call may be very different from the RPM on say the network to which she subscribes or generated by her particular retail plan. Indeed, the caller may not even know what the call will cost. Therefore an empirical demand curve, relating price and quantity demanded, may not be adequately defined.

¹⁹ See for example European Commission (2009a).

- Thirdly, from both consumer and supplier positions the transaction of buying and selling a call is not like buying and selling a banana. The consumer must first invest in a handset and acquire a SIM card. Then she is faced with a myriad of rate plans: pre-paid, post-paid, different combinations of subscription and per minute charges, different on-net/off-net differentials. The different rate plans appear as manifestation of competition but in a world of asymmetric competing networks with high termination rates may be the result of incentives for individual networks to compete for subscribers that have little net benefit to consumers as a whole.
- Not only is there a problem with defining and observing the 'price' of mobile telephony that faces consumers, there is also a problem in defining the quantity side i.e. supply or consumption. Thus the standard measure is minutes of use (MOU) per subscriber. But what does this measure mean? MOU is a product of the number of calls and the average duration of a call and the mix is surely a function of the price structure e.g. flat rate vs. charging by the minute. Secondly, how to count calls received? If receiving a call generates utility and consumers are willing to pay for it then receiving a call should be counted as consumption.
- Voice telephony is what is known as a *two-sided market*. In other words each call involves two parties with each with very likely different valuations of the call. This means that efficient pricing may on the surface look quite different from what would be regarded as efficient in an ordinary market, e.g. price discrimination, which ordinarily maybe regarded as undesirable may be consistent with efficiency in a two sided market.
- The issues discussed above apply mainly at the regulatory level when trying to measure the impact of mobile call termination on consumer welfare. There are also problems at the level of the individual consumer. In particular in terms of information asymmetries, which makes it difficult for the consumer to weigh the benefits of telephone call against its cost. In the case of mobile phone telephony, information asymmetries arise when consumers might not know (or might find it very hard do find out) the price of a telephone call. This is particular the case when operators apply termination-based price discrimination at the retail level. Under such a regime, to know the price of a call, the caller has to know whether the receiver is on-net or off-net. Drawing on the findings of Gans and King (2000, 2005), analyzing fixed-to-mobile, it seems reasonable to argue that in the case of off-net mobile-to-mobile with customers ignorant as to the network they are calling, operators under a Calling Party Pays (CPP) regime will be less constrained by the demand side forces when setting the price of an off-net call. Hence generating prices for off-net calls that are too high in terms of social welfare.

In principle, it would be possible to model all these effects in a single comprehensive model. With sufficient information about parameters it would then be possible to balance the benefits and costs of alternative arrangements. To date nothing like this has been done and the relative strength of various effects has to rely on judgment.

4.1. The channels of impact of MTRs

Ultimately, the social benefit of a mobile termination regime depends on how it affects the choices facing consumers and the ways they have to pay for those choices i.e. it depends on both the structure and level of the tariffs they pay. Some of the complexity of the channels of impact from the mobile termination rate (MTR) to consumers is illustrated in the schematic diagram below:

A key feature of MTRs is that they generate both outlays (costs) and revenues (income) for operators. The termination charges could either be asymmetric or symmetric. When termination charges at the wholesale level are asymmetric as between fixed and mobile and as between off-net/on-net mobile-to-mobile (under certain regimes), then this tends, at the retail level, to result in asymmetric charges for calls between fixed and mobile networks and on-net/off-net (as well as an imbalance in revenues). However, even when termination charges between networks are symmetric (symmetry being imposed by the regulator – cf. Table 1), so that no major transfers occur between the networks, the dual character of termination charges may influence the commercial behaviour (i.e. at the retail level) of the operators in rather significant ways.

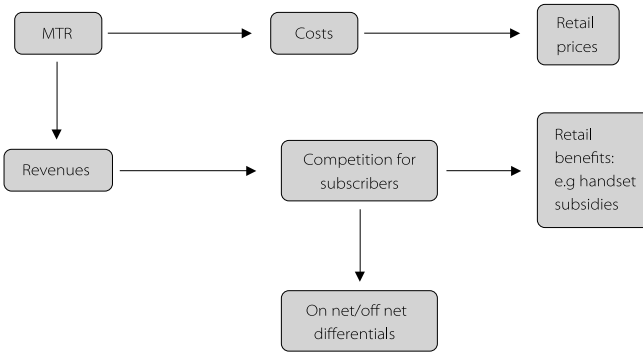


Figure 4. Channels of impact of MTR

4.2. Relationship between MTR and retail prices

The first channel of impact illustrated in Figure 4 is from MTR to costs and then to prices. Thus higher MTRs lead to higher costs for the originating operator and these can be expected to be passed on in higher retail prices for mobile calls. What is the evidence for this?

The most recent investigation of this has been by Pesendorfer and Verone (2009) in a study undertaken for Ofcom, who find that “the statistical relationship between alternative measures of price and level of MTRs ... is weak. This weak relationship emerges for all price measures.” (p. 28). At

the same time Pesendorfer and Verone conclude that they found “a small but statistically significant coefficient for the level of MTRs in all specifications we examined. This finding would be consistent with the argument that operators increase prices to recover higher MTRs.” (p. 28). The model used by the European Commission (2009) study assumes a 33% pass through of MTR changes to retail prices.

In the preparation of this report we made some very rough in-house calculation of the relationship between MTR and retail prices, where the latter is measured as the cost of pre-defined OECD usage basket. We find positive but weak relationships, the strongest of which is illustrated in Figure 5 below²⁰.

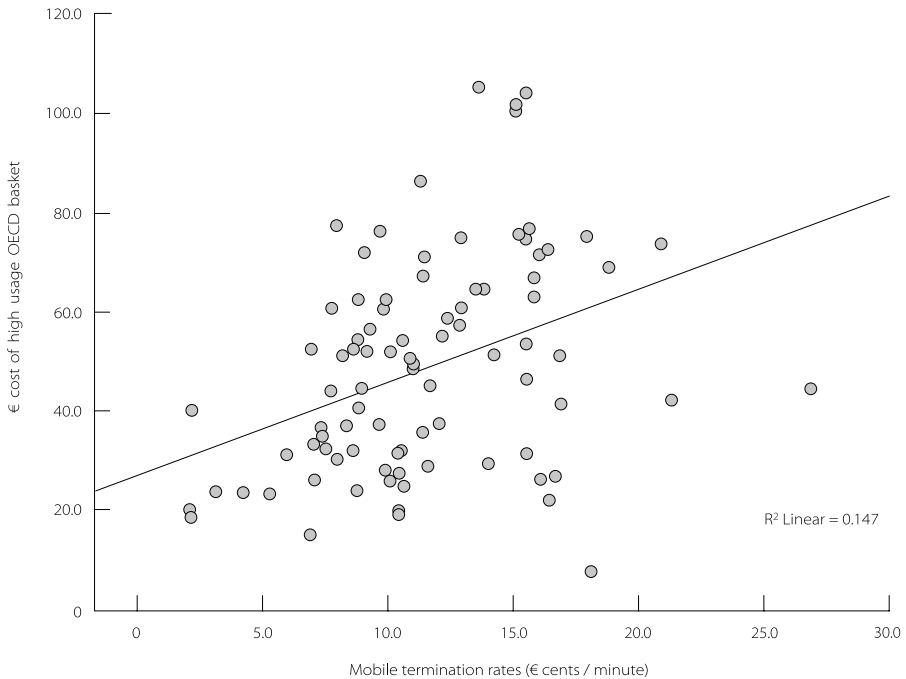


Figure 5. The correlation between MTRs and retail prices

Overall interpretation is difficult. It is clear that the pass through of the costs associated with termination is significantly less than 100%. Perhaps this is simply a reflection of the fact that at the retail level the market structure of mobile telephony is typically oligopolistic and in an oligopolistic market cost changes are only partially passed on. Nevertheless, lower termination rates can be expected to result in lower retail call charges (for the calling party).

²⁰ For medium and low user baskets the relationships are much weaker.

4.3. The revenue channel of impact

Analysis of the revenue channel of impact is more challenging. As noted above, under the Calling Party Network Pays (CPNP) regime, as well as representing a cost to a mobile network operator (MNO) for calls originating in its network, the termination charge also generates revenues from calls originating in other networks. Even if outlays and revenues are balanced, so that an operator's net cash flow from termination is approximately zero, the presence of termination revenues per se is likely to influence the retail options offered by the MNO. This channel of impact of mobile termination rates on consumers has at least two sub-channels. One operates through what is known as the *waterbed effect* and the other operates through the incentive created for MNOs to price discriminate between on-net and off-net calls. The incentive to discriminate between on-net and off-net is also affected by the presence of call externalities.

The waterbed effect

The waterbed effect is defined by Schiff (2008) as a situation where “regulation of one of the prices charged by a multiproduct firm will result in a change of the firm's other unregulated price(s)”. In the context of mobile telephony, termination is one of several services provided by an MNO, and if regulation results in a lower termination charge the waterbed effect implies that MNOs will seek to recoup the lost revenues by raising unregulated prices. The general thrust of this argument is that if there is a waterbed effect then what consumers gain through reduced prices as a result of lower termination charges they will lose from higher subscriptions or reduced handset subsidies²¹. However, consumer welfare may also be affected by the changed structure of prices even though the total outlay (revenue received by the operator) is unchanged. Figure 6 illustrates this by showing how a revenue neutral change in prices can influence consumer welfare.

In Figure 6 we suppose that there is a positive (regulated) off-net termination charge which is seen as a cost and is 100% passed on and results in a retail tariff of T_0 and a call volume of Q_0 and hence implies a call revenue of $T_0 \times Q_0$. We assume that the termination charge is symmetric and calls with other networks are balanced so that termination also generates revenues equal to $(T_0 \times Q_0)$.

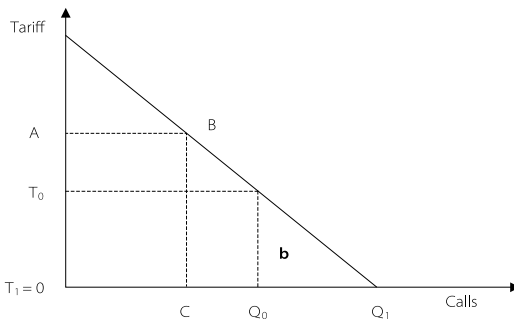


Figure 6. Subscription vs. variable charge

²¹ Within the group of consumers the effect might differ. Some consumers might be net gainers whereas others might be net losers.

Suppose now the regulator imposes a termination charge of zero (as would be the case with BAK) but because of the waterbed effect the subscription fee is increased to fully compensate for the loss of revenues. Because the operator now has no termination charges to pay, the retail tariff will be reduced to $T_1 = 0$ ²² and hence consumers will increase the calls they make to Q_1 . The subscription fee is given by the area of the quadrangle 0ABC in Figure 6 which by assumption is equal to the foregone termination revenues ($T_0 \times Q_0$). At the same time the MNO no longer pays termination charges and no longer collects call revenues from consumers to cover these costs.

Thus the changes are revenue neutral for the MNO i.e. there is a 100% waterbed effect, Consumers also pay the same but in a different way – the *structure* of retail charges has changed in such a way as to encourage more usage. This higher usage generates an extra consumer benefit equal to the area **b** in Figure 6. Thus the overall effect of the change is welfare or efficiency enhancing.

On-net/off-net charge differentials

The existence of large differentials between on-net and off-net call charges (i.e. at retail level) in CPNP regimes is well known, e.g. Birke and Swann (2006) report that in the early 2000s the price of off-net calls in the UK was nearly five times that of on-net calls. Getting a good picture of the current state of this relationship is difficult because of the growing prevalence of plans offered by MNOs that provide different numbers of any-network, any-time allowances or buckets. Moreover in the UK, for example, some MNOs no longer apply an on-net/off-net differential even when the allowances are used up.

However, others do: thus Orange charges 12 ppm (pence per minute) for on-net calls and 35 ppm for off-net calls once the allowance is used up, giving an on-/off-net price differential of 23 ppm; or Vodafone, on its lowest price tariffs, which provide 100 or 250 minutes of any-network calls per month, charges 18 ppm for on-net and 40 ppm for off-net calls outside the allowance giving an on/off-net price differential: of 22 ppm. For tariffs higher than £30 per month, these charges fall to 15 ppm and 35 ppm giving a slightly lower on-/off-net price differential of 20 ppm.^{23 24}

The result of on-net/off-net charge differentials is considerable imbalances in on-net/off-net calls. This is illustrated in Table 3 for a selection of European countries.

Assuming equal costs associated with on-net and off-net calls and charges that reflect costs and assuming a balanced calling pattern i.e. each subscriber calls every other with the same probability, the we would not expect to observe any difference in the volume of on-net and off-net calls, whereas the observed pattern of calls suggests that in many European countries on-net calls dominate off-net by a factor of 2 to 3. This presumably is the result of the pricing structure – in particular the on-net/off-net tariff differential.

The on-net/off-net charge differential and its causes and effects have been somewhat ignored by regulators as a manifestation of economic inefficiency, perhaps because the mobile retail markets and charges are unregulated. However, if the termination charge is set in excess of marginal cost

²² For simplicity we assume there are no other call costs. In practice, this would not be the case but the force of the example remains.

²³ The data are from Ofcom (2009).

²⁴ Additionally in plans with unlimited on-net calls, the effective differential is even greater.

(as is the case of termination charges based on LRIC+) off-net calls will be priced higher than on-net and there is a direct economic inefficiency effect in the sense that too few off-net calls will be made. This is a standard static form of market failure²⁵. On top of this, the revenues generated by terminating calls from other networks creates a direct incentive for, possibly excessive, competition for subscribers, including, for example, cross-subsidisation of low usage subscribers because their calls deficit yields positive net termination revenues.

Country/year	2003	2004	2005	2006	2007	2008
Austria	2.26	1.92	1.86	1.97		
Belgium	2.70	2.18	2.12	2.35		
Cyprus			7.16	3.37		
Czech Rep.	1.68	1.58	1.69	1.77		
Germany	1.74	1.50	1.60	2.14		
Estonia	1.50	1.83	1.52	1.63		
Spain	1.94	1.95	2.16	2.16	2.25	2.03
France	1.71	1.74	2.15	2.27	2.09	1.96
Hungary	1.85	1.69	1.73	1.92	2.09	2.23
Lithuania	2.13	2.38	2.47	3.15	3.21	3.16
Poland	1.45	1.37	1.58	2.17		
Portugal	2.78	3.20	2.96	3.00	3.02	3.29
Sweden	1.80	1.70	2.37	2.10	1.99	1.99
UK	1.86	1.55	1.33	1.18	1.15	

Table 3. Ratio of on-net/off-net call duration (excluding calls to fixed networks)

Source: Eurostat and own calculations

However, when call externalities are present MNOs have a further incentive to set on-net/off-net price differentials. This phenomenon has been described by Laffont, Rey and Tirole (1998) as *tariff mediated externalities*. Harbord and Pagnozzi (2008) in a survey article argue that on-net/off-net price differentials “make larger networks more attractive to consumers than smaller networks. When on-net calls are priced below off-net calls, ceteris paribus, subscribers to large networks experience lower average call charges than subscribers to smaller networks, since more of their calls are made on-net. This makes larger networks more attractive and places smaller networks at a competitive disadvantage.”

In other words a network has a strategic incentive to reduce the number of calls that subscribers on other networks receive and that this incentive is present even when termination charges are priced at marginal cost. Moreover according to Harbord and Pagnozzi the theoretical literature suggests that:

- Large networks are likely to charge higher off-net prices, and create higher on-net/off-net price differentials, than smaller networks. This has the effect of reducing the attractiveness of smaller networks, since subscribers to a large network can be expected to make proportionately more on-net calls than the customers of a smaller network (or subscribers to small networks will make

²⁵ If there is also a call externality i.e. the receiver benefits from the call, then the socially efficient call charge both off-net and on-net should reflect this i.e. charges should be below marginal cost to reflect the externality.

relatively more off-net calls). This implies that subscribers to smaller networks will experience higher average call charges, placing the smaller network at a competitive disadvantage, simply because a network is smaller and not because of any inherent cost differences.

- Above marginal cost mobile-to-mobile termination charges can lead to permanent “access deficits” for smaller networks. Call externalities reinforce this effect, since when large networks set high off-net prices, subscribers of a smaller network will also receive relatively few calls.

5. Estimates of the consumer benefit impact of changes in MTRs

Rather few estimates of the costs and benefits of changing MTRs are available. Perhaps this is unsurprising because modeling the effects numerically is rather challenging, requiring many assumptions in order to come up with results.

According to Littlechild (2006):

“In 1998 the MMC²⁶ made no calculation of the benefits of introducing a price control to reduce termination charges ... more recently, however, regulatory authorities have sought to quantify the costs and benefits involved, or have been obliged to do so. This requires a prediction of what prices and the allocation of resources would be, both with and without the proposed price control. It is then necessary to assess what the costs and values of the resulting inputs and outputs would be. This is by no means a straightforward task and for the most part not one that regulators of other utility sectors have attempted. Nor have the telecom regulators attempted to quantify the so-called ‘dynamic’ efficiencies associated with competition and innovation, as opposed to the income transfers and the ‘static’ efficiencies associated with improved resource misallocation.”

Table 4 below summarises the estimated impact of the price control made by regulators in the early 2000s in UK, Australia and New Zealand as reported by Littlechild (2006).

	UK (2003)	Australia (2004)	New Zealand (2005)
Annual wealth transfer	950-1267m	1154m	16-17m
Annual welfare gain	205- 443m	220m	5-7m
Annual welfare gain per capita	3-7	11	1-2

Table 4. Wealth transfers and welfare gains from termination charge price controls (converted to USD)

Source: Littlechild (2006)

²⁶ The UK Monopolies and Mergers Commission.

It is clear, as put by Littlechild: "...that net welfare gains are up to an order of magnitude less than wealth transfers. More surprising, perhaps, is the rather small size of all these magnitudes, particularly on a per capita basis" Littlechild also notes that the MNOs "challenged almost all the calculations in the price control reviews, heavily disputed even the low figures on welfare gain". A particular point of criticism was that that "the regulators had failed to give adequate weight to the waterbed effect, with its implied increase in subscription charges and reduction in number of subscribers."

The most recent exercise of this kind has been by the Commission in European Commission (2009a) where estimates are offered for the transfer and welfare impact of moving from the current regulatory regime to one that effectively bases termination rates on long run marginal cost. The analysis assumes that on average termination rates would fall by an extra 30% over 2009–2012 as compared with the current regime. The Commission supposes that this will result in an extra 8.1% reduction in mobile retail prices. The analysis does not include waterbed effects, but the pass-through of the termination rate reduction is assumed to be quite conservative at 0.33. A quite modest price elasticity is also assumed generating an extra cumulative increase in mobile traffic. The net cumulative over 2009–2012 welfare effect of this is estimated at 1 billion euro and given an EU population of 500 million this implies a per capita welfare gain of 2 euro. This is really a rather modest figure and suggests that decreasing returns to tighter cost based charge regulation may already be observed.

The European Regulators Group have attempted to use data on comparative tariffs and traffic in CPNP countries and BAK countries to define a 'demand curve' and to use this to estimate the welfare effect of a move from CPNP to BAK. The data is illustrated in Figure 7.

There are data problems with such an approach since the most commonly used proxies (see for example Ofcom, 2009) for price and quantity are revenue per minute (RPM) and minutes of use (MOU) which come from Merrill Lynch. It is well known that in Merrill Lynch data MOU are overstated in BAK countries and revenues are overstated in CPNP countries²⁷. The data in Figure 7 use an Ofcom adjustment to correct for these problems. Even so, the price proxy in CPNP countries is about twice as high as in BAK countries. On the other hand usage, as proxied by adjusted MOU, in BAK countries is approximately double that in CPNP countries. If this can be interpreted as a movement along a demand curve it would imply a price elasticity of -2 (as seen from Figure 7 where on average a 50 per cent price difference is associated with a doubling of traffic).

Assuming that adoption of BAK in Europe would shift Europe down this 'demand curve' to where BAK countries are this would generate an annual net welfare gain of 25 billion euro or 50 euro per capita²⁸.

²⁷ Revenues may be overstated because some operators include non service revenues in their average revenue per user figures. See Ofcom (2009), Annex 5, p. 4.

²⁸ This is based on a current MOU of 1000 billion, a fall in price (RPM) from 10 to 5 euro cents and a consequent doubling of usage.

revenue per minute transferred for \$ to euro at 1.51 exchange rate

MoU versus RPM (Merill Lynch 2008 Q3 data)

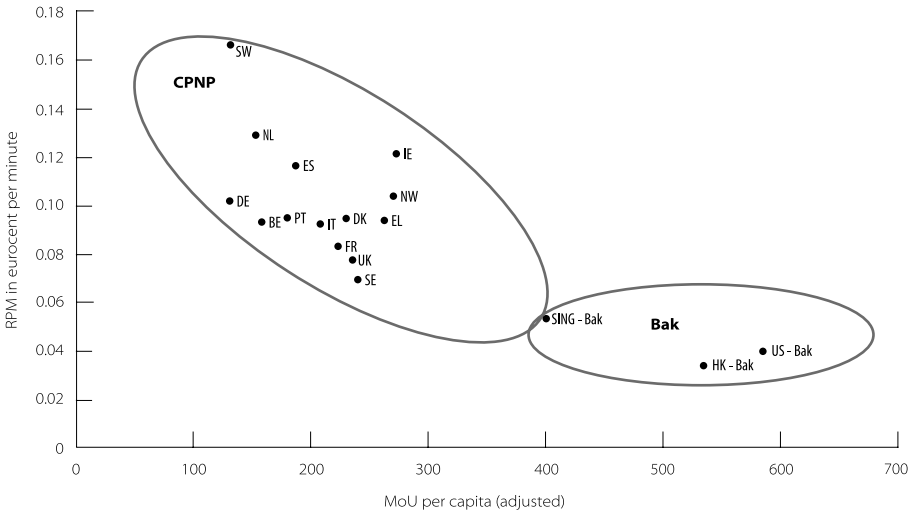


Figure 7. An aggregate demand curve for mobile calls using adjusted RPM and MOU as proxies for price and quantity

This is a serious gain in welfare, but relies on a number of crucial assumptions. One is that the curve implicit in Figure 7 can really be interpreted as a demand curve. This is highly problematic since as has already been noted RPM does not represent the actual price of any actual call. A second factor is the assumption that differences observed as between RPM and MOU are the result of differences in the termination regime and hence that if BAK were adopted in Europe we would observe the same outcome as in the US or Hong Kong. Clearly, many other factors may be at work in generating the observed differences.

Even the ERG does not fully believe its own calculations:

“The ERG does not claim the effect of BAK in Europe would have this huge effect. However, the calculation does illustrate the size of the difference between the European and US [markets] ... [where] the latter is generally considered superior in terms of welfare. The ERG assesses that these results indicate it is at least highly likely that the European market will move in the direction of higher usage and lower price per minute by moving to BAK. This would result in ... higher welfare”. (ERG (2009) pp 39–40)

6. Cost based charge regulation vs. mandated bill and keep

This section attempts to provide a check-list of the pros and cons of the current system of cost based price regulation as compared with a move to mandated bill and keep. The first point to make is that neither regime is likely to result in an unambiguous 'first best' efficiency optimum. By definition, BAK implies that termination charges cannot reflect termination costs as would be the case in a theoretical optimum. On the other hand a cost based price cap is unlikely to exactly reflect marginal costs and in practice is therefore also unlikely to result in the theoretically optimum allocation of resources.

So, which regime is to be preferred? The dimensions along which the regimes may be compared include the following:

- *Economic efficiency*: this includes *static efficiency* i.e. how do the models compare in terms of generating the conditions for MNOs to supply the right (socially desirable) quantity and mix of services at prices that reflect costs and also *dynamic efficiency* i.e. how do the regimes compare in terms of delivering the socially desirable level of investment and innovation.
- *Distributional effects*: how far do the alternative regimes have different impacts on different groups of consumers? And if such differences exist how should they be evaluated?
- *Competition effects*: although competition is not in general an end in itself mobile markets are typically oligopolistic which are sometimes characterized by anti-competitive behaviour.
- *Regulatory effects*: the alternative regimes are likely to have a different regulatory burden.

	Bill and keep	Cost based regulation
Static economic efficiency		
Price and quantity supplied effects of termination monopoly	<p>a) Because, in this regime, termination charges are set at zero and this is reflected in lower average retail prices, usage is likely to be higher. The cross country evidence tends to support this although the evidence is not straightforward to interpret.</p> <p>b) If empirical evidence on RPM and MOU in CPNP countries and BAK can be taken to represent what would happen in a switch to BAK then we could expect a halving of RPM and doubling of MOU, with a net annual welfare gain of 50 euro per capita in the EU.</p>	<p>a) If marginal costs could be accurately identified and tariffs exactly reflected marginal costs this regime would generate a fully efficient allocation of resources. In practice this is unlikely to be the case because marginal costs cannot be precisely identified and regulated prices are in practice based on average cost. This means that in practice some element of efficiency loss remains.</p> <p>b) The measures of consumer welfare gain obtained in empirical studies of cost based regulation have been rather small in per capita terms.</p> <p>c) Regulation has already reduced termination rates considerably and the return to tighter cost-based regulation may be small.</p>
On-net/off net differentials	<p>a) BAK removes the incentive for on-net/off-net differentials that are made simply to capture revenues and which lead to too few off-net calls being made.</p>	<p>a) A termination charge in excess of marginal cost generates an incentive to differentiate between on-net and off-net retail tariffs, which cannot be justified on cost or efficiency grounds.</p> <p>b) The presence of on-net/off-net differentials generates what have been called tariff mediated externalities which can make large networks more attractive than small ones even when there are no inherent cost differences between networks.</p>
Call externalities	<p>a) BAK could be expected to result in RPP at the retail level so that both the calling party and the receiving party pay for a call. This would have the effect of partially internalizing the call externality. The size of the gain from this is unknown because the empirical importance of call externalities is not known.</p>	<p>a) Under CPNP cost based price regulation would be optimal if either the call externality is negligible or if the regulator can correctly identify the value of calls to receivers. There is really no evidence on this and it is inherently difficult so that there is considerable risk of regulatory failure (in the sense of setting the wrong termination rate).</p>
Network externalities	<p>a) Under BAK, tariff mediated externalities are eliminated.</p>	<p>a) High termination charges is probably not an efficient way to finance the subsidy to encourage people to participate.</p>
Penetration rates	<p>a) Penetration rates appear to be lower in BAK regimes. This does not mean they are not optimal. Optimality would require that all potential consumers who are willing to pay the marginal cost of becoming a mobile network subscriber (including a network externality subsidy) in fact do so. We do not know what that rate is. There is certainly no presumption that a higher penetration rate is better than a lower one.</p>	<p>a) Termination rates in excess of marginal cost give operators a possibly excessive incentive to compete for subscribers.</p> <p>b) Penetration rates in excess of 100% suggest that people may have more than one SIM card simply to avoid high off-net charges. This is socially wasteful.</p>

	Bill and keep	Cost based regulation
Dynamic economic efficiency		
Investment incentives	<p>a) The main argument here is the 'hot potato' effect. The argument that this can be eliminated by the definition of the domain of BAK seems persuasive.</p> <p>b) There is no evidence of 'under investment' in BAK regimes.</p>	<p>a) Cost based prices that fully reflect marginal social costs also create the socially correct price signals for investment. To the extent that prices do not fully reflect social costs then the signals for investment may be incorrect.</p>
Distribution effects		
Low usage consumers	<p>a) Pre-pay subscriptions may be less attractive and hence some low usage subscriptions may be discouraged.</p>	<p>a) Low usage subscribers cross-subsidised from termination revenues.</p>
Competition effects		
Small vs. large, mobile vs. fixed	<p>a) Under usage of fixed telephony likely to be reversed.</p>	<p>a) Tariff mediated externalities make small networks unattractive even if they are just as efficient in terms of cost.</p>
Regulatory effects		
	<p>a) In BAK termination rates are automatically set at zero so the direct regulatory burden associated with setting rates is completely avoided.</p> <p>b) There may be regulatory costs associated with managing the transition to BAK.</p> <p>c) Regulation needed to address the hot potato problem.</p>	<p>a) It is widely accepted that cost based regulation imposes a heavy regulatory burden generating costs that have to be covered by taxes and/or fees on the regulated parties. Littlechild (2006, p. 271) suggest that regulatory costs approach USD 300 million per year in Western Europe alone. In addition there are most likely substantial costs associated with litigation.</p>

7. Conclusions

An enormous literature on the issue of mobile termination has emerged. It has been discussed in the academic literature, in official reports and in arguments produced by consultants. Many alternative 'models of termination' have been considered and explored. The aim of this paper has been to provide a synthesis of this literature, at least of its most salient parts. However, this synthesis comes with a 'health warning': the only one certain outcome of all this activity and is that there is no agreement on the best model. Indeed there is not even agreement on how we might decide on

a best model. One reason for this is lack of a clear view of what an unregulated market might look like and hence what the impact of regulation might be. Thus, as Armstrong and Wright (2008) point out, while there is a presumption that the competitive bottleneck arising from a network's monopoly of termination in its own network gives rise to an incentive to raise termination charges above marginal cost, when mobile networks compete there may be an incentive for them to jointly set mutual termination rates below marginal cost so as to soften competition in the retail market. In other words, we do not know where in relation to marginal cost we might observe termination charges in an unregulated market.

The general implication of this is well stated by Ofcom (2009) in their preliminary consultation on the future regulation of wholesale mobile voice termination:

“It is not clear that there is a single termination regime that is better than all the others under all circumstances and for all consumers. ... The economic case for and against each of the candidate regimes is mixed, both in theory and in evidence. We have found that much of the evidence that might help guide our analysis is either difficult to obtain robustly or is open to more than one interpretation.” (pp 34–35)

However, if one is forced to come off the fence, then choosing between two less than perfect alternatives the choice would have to be BAK for perhaps three reasons: it removes some clear inefficiencies present in the current system, it greatly reduces the regulatory burden; and it may be the way forward as convergence is realized.

Finally, with the ongoing convergence of telecommunication to a single IP-format it seems reasonable to suggest that the choice of regulatory regime should be able to handle the new generation network models. However, as discussed in Dodd et al. (2009) NGNs will not remove the problem. WIK-Consult (2008, p. XI) puts it succinctly: “...the migration to IP-based NGNs will not ameliorate the termination monopoly ... until and unless mechanisms emerge to enable more than one telephone service provider to terminate calls to a single telephone number.”

There remains the question of what sort of transition period would be reasonable in changing from to BAK from the current regime. Clearly, the changeover is not going to happen in 2012 so if it happens at all it will be some time after that. There are clearly technical issues to be resolved e.g. how to prevent free riding by providers who have no network of their own, and the regular operators need time to examine how they would adjust their operations under a BAK regime. So something like beginning 2016 might be feasible but a decision would have to be made well in advance e.g. by the end of 2012.

References

Albon, R. and York, R. (2006), Mobile termination: Market power, externalities and their policy implications, *Telecommunications Policy*, 30, 368–384.

Albon, R. and York, R. (2008), Should mobile subscription be subsidised in mature markets?, *Telecommunications Policy*, 32, 294–306.

Analysys Mason (2008) “Case studies of mobile termination regimes in Canada, Hong Kong, Singapore and the USA” Annex 8.1 of Ofcom (2009).

Armstrong and Wright (2008), *Mobile Call Termination*, mimeo, October 2008.

Corrocher, N. and Zirulia, Z. (2009), Me and you and everyone we know: An empirical analysis of local network effects in mobile communications, *Telecommunications Policy*, 33, 68–79.

Dodd, M., Jung, A., Mitchell, B., Paterson, P. and Reynolds P. (2009), Bill-and-keep and the economics of interconnection in next-generation networks, *Telecommunications Policy*, 33, 324–337.

European Regulators Group (2009), ‘Draft common position on next generation networks future charging mechanisms/long term termination issues’.

European Commission (2009a), Working document accompanying the Commission Recommendation on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU: *Implications for Industry, Competition and Consumers* [C(2009) 3359 final].

European Commission (2009b), Commission Recommendation on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU.

European Commission (2009c), Progress report on the single European communications market (14th report), [COM(2009)140 final].

Gans, J.S., King, S.P., and Wright, J. (2005), Wireless Communications, in S.K. Majumdar, I. Vogelsang, and M.E. Cave (eds), *Handbook of Telecommunications Economics*, North-Holland.

Gans, J.S. and King S.P. (2000), Mobile network competition, customer ignorance and fixed-to-mobile call prices, *Information Economics and Policy*, 12, 301–327.

Jeon, D-S., Laffont, J-J, and Tirole, J. (2004), On the “receiver-pays” principle, *RAND Journal of Economics*, 35, no 1, 85–110.

Kennet, M.D. and Ralph E.K. (2007), Efficient interconnection charges and capacity-based pricing, *International Economics and Economic Policy*, 4, 135–158.

Littlechild, S.C. (2006), Mobile termination charges: Calling Party Pays versus Receiving Party Pays, *Telecommunications Policy*, 30, 242–277.

Ofcom (2009), *Wholesale mobile call termination: preliminary consultation on future regulation*.

Post- och telestyrelsen (2002), *Guidelines for the LRIC-bottom up and top-down models*, September 2002.

Ramsey, F. (1927), A Contribution to the Theory of Taxation, *Economic Journal*.

Schiff, A. (2008), The Waterbed Effect and Price Regulation, *Review of Network Economics*.

Vogelsang, I. (2003), Price Regulation of Access to Telecommunications Networks, *Journal of Economic Literature*, September, 830-862.

WIK Consult (2008), *The future of IP interconnection: technical, economic and public policy aspects*, January 29, 2008.

Mobile Termination:

How to Regulate
or Perhaps Not to
Regulate at All?

TeliaSonera Institute

Strēlnieku iela 4a, Riga, LV-1010, Latvia

Baltic International Centre for Economic Policy Studies (BICEPS)

Strēlnieku iela 4a, Riga, LV-1010, Latvia

www.biceps.org

Stockholm School of Economics in Riga

Strēlnieku iela 4a, Riga, LV-1010, Latvia

www.sseriga.edu.lv