Contract No. 48544
Cost Structures in Mobile Networks and their Relationship to Prices

Final Report
for the
European Commission
by
Europe Economics

Europe Economics
Chancery House
53-64 Chancery Lane
London WC2A 1QU
Tel: (+44) (0) 20 7831 4717
Fax: (+44) (0) 20 7831 4515
www.Europe-economics.com

28 November 2001

TABLE OF CONTENTS

EXECUTIVE SUMMARY I

1 INTRODUCTION 1
1.1 Mobile Telephony Markets in the EU 1
1.2 Purpose and Structure of this Report 2

2 REGULATORY ACTIONS AND INITIATIVES 4
2.1 Overview 4
2.2 Notified Operators 9
2.3 Price Controls 10
2.4 Measures to Promote Competition 12
2.5 Identifying Objectives 14

3 COST ELEMENTS AND DRIVERS OF MOBILE NETWORKS 15
3.1 Costing Using Published Accounts 15
3.2 GSM Network Architecture 16
3.3 The Concept of Coverage 21
3.4 Cost Drivers in a Mobile Network 22
3.5 Implications 28

4 MARKETS FOR MOBILE SERVICES 30
4.1 Nature of Retail and Wholesale Markets 31
4.2 Defining Origination and Termination Services 33
4.3 Roaming Services 35
4.4 Summary 36

5 COSTING SERVICES 38
5.1 Costs Characteristics of a Telecommunications Network 38
5.2 Joint Costs in Mobile Telephony 39
5.3 Approaches to Estimation of Costs 41
5.4 Defining Increments 44
5.5 Recommendations if Building a Cost Model 52

6 EFFICIENT COSTS IN A DYNAMIC SETTING 54
6.1 Productive Efficiency in Mobile Telephony 54
6.2 Dimensioning the Network 56
6.3 Attributing Costs in a Dynamic Environment 58
6.4 Regulatory Implications 62

7 COSTS AND PRICES 64
7.1 Pricing with Large Joint Costs 64
7.2 Relationship between Tariff Structures and Costs 70
7.3 Dynamic Profile of Prices 73
7.4 Cross-Country Price Comparisons 74
7.5 Implications for Developing Regulatory Framework 77

8 ROAMING 78
8.1 Charging Arrangements 78
8.2 Network Costs of Roaming 80
8.3 Comparing Retail Roaming Prices and Costs 84

9 REGULATING MOBILE COSTS AND PRICES 86
9.1 Introduction 86
9.2 How Mobile is Different From Fixed 86
9.3 Weighing up Different Regulatory Strategies 87
9.4 Non-discrimination 88
9.5 The Value of Regulatory Accounts 89
9.6 Building Cost Models 90

APPENDIX 1: AVERAGE COST PER MINUTE CALCULATIONS 88
A1.1 Wholesale vs. Retail Costs 88
A1.2 Worked Example of LRAIC per minute 90

APPENDIX 2: OPERATIONS PROVIDED BY A GSM NETWORK 100
A2.1 Location Updating and Authentication 100
A2.2 Voice Calls 101
A2.3 Billing 107
A2.4 Other Operations 107
A2.5 Roaming 109

APPENDIX 3: ABBREVIATIONS USED 114

APPENDIX 4: ACKNOWLEDGEMENTS 116

APPENDIX 5: SOURCES FOR PRICE INFORMATION 117

EXECUTIVE SUMMARY

Introduction

This report is the main output of a project undertaken by Europe Economics for the Information Society Directorate General of the European Commission (Commission). The main task of the study was an examination of the cost structures of mobile networks and their relationship to prices. This was done with a view to producing a report that provides the foundations for policy analysis and initiatives, and to make suggestions for possible recommendation or guidelines on costing and pricing in mobile networks. This required an examination of the cost structure of mobile networks and their relationship to prices.

The European Commission's motivation for looking at costs and prices in mobile telephony stems, in part, from concerns that whilst the industry is generally competitive, there have been concerns raised about the competitiveness of the supply of a few services. For example, there is a perception that the incentives are not in place for competition in setting call termination rates. Similarly, the charges mobile operators levy on consumers who roam abroad appear to bear no relationship to the prices consumers pay for similar services on their own network.

Regulatory Actions

Our starting point is a review of the regulatory environment around Europe. The main findings are summarised in Section 2. More details, on a country-by-country basis, are provided in the Annex to the report.

There are two markets in which mobile operators can be judged to have significant market power (SMP) – the mobile market and the market for interconnection. In 13 of the member countries of the EU, at least one operator has been found to have SMP in the mobile market; eight of these countries have found one or more operator to have SMP in the interconnection market.

Operators designated with SMP in the interconnection market, must observe cost orientation. Price caps on termination rates for fixed-to-mobile calls have been introduced in eight member states. Some of these countries also regulate the price for mobile-to-mobile call termination. Most regulators that have imposed price caps have to date relied on either best-practice rates or fully-allocated cost models.

In addition, NRAs can impose other regulations on operators, including those not designated with SMP, to encourage competition, provided that the regulations are proportionate and based on objective criteria. Thus, for example, NRAs may introduce polices concerned with access for MVNOs, national roaming rights for 2G and 3G MNOs and site sharing arrangements between MNOs.

The approaches taken have differed by regulator. This need not imply that some regulators have adopted inappropriate policies. Instead, the situation may differ by countries, or the goals of the NRAs may differ. If looking at costs and their relationship to prices, a sensible first step is to establish the goals that the regulator wishes to achieve. How to think about cost structures will partly depend on the reason for looking at the costs.

Costs of a Network

A review of the overall competitiveness of the mobile industry may just be concerned with comparing total costs with total revenues. Guidelines may be necessary to determine the appropriate way to calculate the economic costs, as opposed to the reported accounting costs. Differences may arise because of the way depreciation is estimated, or because of the difference between current and historic costs. But NRAs have experience addressing such problems regulating fixed network operators.

If the focus is on the costs of individual services, such as call termination or roaming, a more detailed understanding of the costs of a mobile network will be necessary. (For reasons explained in the report, calculating the average cost by dividing total costs by total minutes produces a number that is not helpful for formulating a regulatory response.) Mobile networks contain a variety of different elements, only some of which have direct correspondence with elements in fixed networks. Base transceiver stations, backhaul, base station controllers, mobile switching centres, transmission links, and spectrum are all important elements in a mobile network.

Many of the differences between mobile and fixed networks when thinking about costs arise because of the fact that mobile network customers can access the network from anywhere, rather than from a defined point on the network. Mobile networks provide coverage. There are a variety of ways to define coverage. We consider that for thinking about costs, defining coverage as the capability to make a single call from any point of the network at a point in time is a sensible approach. Other possible definitions are discussed.

Unlike subscribers and call minutes, where an individual can be identified as responsible for a certain proportion of the total volume of the cost driver, this is not possible for the coverage. No-one is individually responsible for a mobile network providing coverage in Berlin. An NRA wishing to see prices set according to the principle of cost causality, popular for fixed-network regulation, will need to determine how such a principle can be adapted for mobile networks to take into account the fact that no subscriber causes coverage costs individually. The cost drivers for network elements will differ. The major cost drivers are the number of subscribers, the volume of traffic (call attempts and call minutes) and the geographical area covered by the network. For many of the elements, there is more than one cost driver. Most elements are influenced both by the geographical area to be covered and the volume of traffic. This complicates attempts to make general comments about the sensitivity of costs to changes in a cost driver. For example, doubling the size of the geographical area covered by a network that carries almost no traffic will almost double the cost of the radio network (the costs of sites and BTSs will double). Yet the costs of two networks that both carry large volumes of traffic may be the same, even if one covers an area twice the size of the other.

Other factors that will influence the cost of mobile network, beyond the three main cost drivers, are differences in terrain, availability of spectrum, and the ease of obtaining planning permission.

All of these factors may differ by country, although trying to quantify the impact for the purpose of cross-country comparisons is complicated by the fact that their impact will depend on the size of the main cost drivers. Terrain can affect the costs of coverage, most notably in the case of mountains — a mountainous region may require many more sites to cover the same area as a flat region. Network operators with more spectrum will be able to reduce the amount of cell splitting that is necessary, reducing the costs of sites and BTSs. Differences in planning restrictions, including differences in health and safety regulations, may increase the cost of providing coverage.

Markets and Services

In mobile networks the distinction between wholesale and retail markets may be a useful one to make. The competitiveness of the retail market inevitably depends in part on the nature of the wholesale market. The scarcity of spectrum means that entry is restricted in the wholesale market. There is a limit to the number of competitors. The possibility exists for intermediaries to compete with MNOs in the retail market. The intermediaries may be mobile virtual network operators; they may be airtime resellers. Competition in the retail market will involve competing tariff structures, quality and ranges of services.

To cost individual services, and compare these to the prices of those services, requires as a first step a definition of the services. This seemingly academic exercise is important. For mobile networks, there is more than one way to define the services in ways that appear reasonable. The main controversy surrounds what services include coverage.

Is coverage a separate service that consumers purchase separately or is coverage part of a bundled service that consumers pay for when they pay for call origination or termination? In a fixed network, the services more easily divide into access network services and core network services. In a mobile network, it is unclear whether coverage is part of an access service or part of a conveyance service. We argue that to define coverage either way is defensible. But the choice has implications for the costs calculated for the different services. We favour including coverage as part of the service a customer purchases when making a call.

One practical advantage of our approach is that the costs estimated for call termination will be the same as the costs estimated for allowing a roaming customer to terminate a call, a scenario where it would be harder to argue that the network permitting the roaming has already received a payment for the costs of the coverage provided. Thus a model built to look at wholesale costs would not have to distinguish between roamed and non-roamed traffic. If looking at retail roaming costs, then it would be necessary to include retail costs such as billing and advertising. These are costs incurred to provide the service.

Costing Services

It is important to be clear about what is meant by the cost of a service. In economics there are a number of cost concepts: incremental, stand-alone, average, marginal, total, joint, etc. Calculating the cost of a service needs not just a definition of the service, but also a clear understanding of which cost it is that is being estimated. It is also important to understand that cost and price are two different concepts.

For a given cost concept, the per minute cost of call termination and call origination is likely to be similar. But for mobile networks there will be large joint costs between services. We recommend that any cost model that is built should clearly identify such costs. Because joint costs are large, it is impossible to identify a unique value for the cost of different services. The incremental and stand-alone costs of call termination will be very different. Ditto for call origination. The differences will be more significant than is the case for fixed telephony. It will not be unusual for the stand-alone cost of a service to be over 100 per cent higher than its incremental cost. For networks with low market penetration, it may even by the case that the incremental cost of voice traffic is zero.

FL-LRIC methodologies are increasingly being used to provide estimates of the costs of services in fixed networks. Such methods will be useful in mobile settings as a means of identifying joint costs. The definition of the increments should not be motivated by a desire to minimise the measured size of the joint costs. The size of the joint costs is invariant to the methodology used; some methodologies may merely be less transparent and fail to measure the joint costs. We recommend that there should be three increments — subscribers, traffic and coverage — corresponding to the three main cost drivers. The incremental cost of traffic will be the costs on top of providing coverage and serving subscribers, that the network incurs to be able to carry the traffic load. A LRAIC approach could be used to allocate the incremental costs of traffic to the different services within the increment. Attempting to calculate the LRIC of the services will produce results that depend on the sequence in which the incremental costs are calculated. The model should estimate the stand-alone costs of coverage. These costs will almost all be joint with the traffic increment. They also represent the bulk of the joint costs between call origination and call termination.

Dynamic Considerations

In fixed networks one of the concerns has been that incumbent operators with legacy networks should not receive compensation for inefficiently incurred costs. Such concerns are in our opinion less likely to be a concern in mobile networks. Assuming there is competition between MNOs in the provision of some services, competitive pressures will force the operators to be efficient in the provision of these services. Since different services use the same network elements, this should mean that productive efficiency holds for all services.

Competition between network operators should not only ensure that the operators are reasonably efficient in the production of a service. It may also mean that they provide different services, or different qualities of service. In the case of service quality, improving the quality of call origination will improve the quality of call termination. The regulator will have to make a judgement on what quality of service the calling party should pay for if attempting to cost the service of call origination for the purposes of setting prices. There is considerable merit in setting this with reference to the quality that a caller to the network actually receives rather than some notion of the efficient quality of service that an operator should provide.

The most obvious example of new services concerns the rollout of 3G networks, with the added features that this technology promises to offer. Costs incurred to provide these new services do not represent the costs of providing current services. The cost of terminating a voice call is independent of the cost of a 3G-spectrum license. However, in many instances it will be difficult to isolate the costs of current and future services. The network topology of 3G networks will likely differ to that for 2G services. An efficient MNO will build its network with a view to migrating to 3G technologies. It will not seek to minimise costs at a point in time, but rather minimise costs in the long run.

The regulator will need to trade-off between theoretical correctness and what is practical. In assessing the trade-off the goals of the regulator will be important. For most exercises, we would, for example, recommend that the sites of the network operator are taken as given (a scorched-node assumption), even if some of the sites are only necessary because of the plans to move to a 3G network. Similarly, in many instances the company's cost of capital should be used as an estimate for the cost of capital of a service.

Costs and Prices

The main result that emerges from our study is that costs alone are not sufficient to explain observed price differences. Nor is it satisfactory to focus solely on costs if the intention is to regulate prices. The presence of joint costs, mainly between coverage and traffic, means that demand conditions are important.

If the regulator's goal is for prices to correspond to those in a competitive market, then prices should be such that the stand-alone costs of each service are not exceeded by the revenue earned on that service. Moreover, total revenues should not exceed total costs. If these conditions hold, then the outcome is consistent with a competitive industry.

Alternatively, the regulator may wish to observe prices that realise allocative efficiency, subject to the constraint that MNOs should be able to break-even. In this case Ramey prices may be appropriate — the services for which demand is least price sensitive should be the ones from which most of the joint costs are recovered. Competing firms would not necessarily choose the efficient Ramsey prices. If competition is greater for call origination than for call termination, the unregulated prices of call origination would be lower than the industry-level Ramsey price and call origination prices would be too high. To realise allocative efficiency, the regulator would need to regulate the prices of all the operators, not just an individual operator. Moreover, to do this properly would require estimation of demand elasticities. Obtaining reliable estimates of demand elasticities for different mobile services is not easy and may be impractical in a market with demand changing. But since joint costs are large in mobile telephony, even if the exact elasticities cannot be calculated, good regulation should give some thought to the likely differences in elasticity between different services. Using equi-proportionate mark-ups to set cost-based price caps, without regard to demand conditions, may reduce rather than increase allocative efficiency.

The tariff structures offered in the retail market for mobile telephony provide evidence that competitive markets do not necessarily result in cost-based pricing. Operators often offer a range of different tariff structures, allowing customers to self-select the most suitable plan given their calling needs. The customer may pay close to incremental cost for those services (s)he most values, while paying closer to stand-alone cost for less important services. In this way, the operator is able to attract more customers and also earn the mark-up on some services that is necessary if joint costs are to be recovered. It is neither efficient nor desirable for all customers to face the same uniform tariff for each service the operator provides.

Another complication with comparing prices to costs is that firms are operating in a dynamic environment. Their pricing decisions, as with their investment decisions, will be set to optimise profits over a period of time rather than at a single point in time. Consequently, there will be some dates when costs exceed prices and other periods when prices exceed costs. To insist on prices equal to costs at all dates would be to deter investments in new technologies. The prospect of earning profits in the future on an investment is necessary for the firm to invest in the first place and risk making a loss.

Because prices depend on both supply (costs) and demand conditions in a market, there is a limit to what can be learned from undertaking cross-country comparisons. Comparing the price of a service with the price charged in other countries for that service will not provide information on the relative efficiency or competitiveness of that countries mobile industry. The joint costs may be recovered differently in different countries. Even looking at prices of all services and trying to draw conclusions is limited by dynamic considerations — networks and competition may evolve differently in different countries.

Roaming

The costs, however defined, of providing roaming services in the wholesale market are not too dissimilar to the costs of providing the equivalent services to the networks own customers. For call origination, there is almost no difference in terms of costs, between roamed and non-roamed calls. The differences are larger for terminating calls, but these differences are not enough to explain the large price differences that have been reported.

But for the same reasons as outlined above, the focus on costs alone to explain price differences is mistaken. Demand conditions are also important. For example, if the roaming customer is typically using a business phone we might expect reduced price sensitivity compared to residential users.[1] Comparing costs with prices should be confined to checking that the stand-alone cost of the network services, including the additional cost of facilities enabling roaming services, is less than its price. If this is not the case, regulators will need to consider whether competition can be increased in the market or whether a price cap is warranted. Other examinations about the prices of roaming services will need to consider supply and demand.

For retail roaming, the IOT represents most of the network costs associated with a call. However, the operator will also incur costs associated with retailing, such as billing and advertising. Many of these costs will be joint across a range of retail services, including a roaming service. So

1 NRAs and the EC are increasingly receiving complaints from consumers travelling abroad about what they perceive as high charges for roaming as compared with domestic mobile charges.

demand conditions will matter for the price of the retail services, if considering what the "competitive" price should be or what an "efficient" price would be.

Summary

To summarise, there are important differences between fixed and mobile networks. *Applying the same approach to regulate mobile networks as was applied to fixed networks may not be appropriate*

- There should be less of a concern over inefficient network infrastructure — regulation should not target productive efficiency in mobile networks without good reason.

- The provision of coverage is a substantial cost in mobile telephony that cannot be attributed to individual users or services by cost causality. There is no parallel to this in fixed networks.

- Because of the large joint costs in mobile telephony it is even more important to take a sophisticated view of the relationship that arises between costs and prices in competitive markets — there are many different ways that prices could be set that are consistent with the workings of a competitive industry.

We recommend that before initiating any cost analysis a regulator should clearly identify what goals they wish to achieve, and the relative importance of these different objectives. *Regulatory action to control prices might be justified if*

- The revenues that operators participating in the market earn from a service are persistently above the stand-alone cost of the service; or

- There is a strong case that allocative efficiency can be improved or distributional goals realised by rebalancing prices between different services or users and the full implications of this intervention have been thought through.

Intervention should not be justified solely because prices for one service are different from prices for another technically similar service; nor because prices for a service in one country appear different to prices for the same service in another country.

Where intervention is considered necessary, *regulatory transparency is desirable*. If costing services, the regulator will need to define clearly the services being considered. A good cost model will explicitly identify any joint costs, and justify the rules used to allocate these. The detailed recommendations included in this report are intended to help regulators in understanding key issues if contemplating building a cost model for the control of prices of mobile services.

1 INTRODUCTION

1.1 Mobile Telephony Markets in the EU

Two developments characterise mobile telephony in the EU in recent years. The first is the rapid growth of the proportion of the population owning a mobile phone. Second, as the mobile market has been liberalised more mobile network operators (MNOs) have been licensed and have begun operations. Table 1.1 below provides evidence of these two developments. It shows rising penetration rates and most markets becoming less concentrated between 1997 and 2000. The market penetration rate is the percentage of the population owning a mobile phone, whilst concentration in terms of the numbers of subscribers has been measured by a Herfindahl concentration index (calculated as a sum of the squares of the firms' shares of market subscribers).[2]

Table 1.1: Developments in European Mobile Markets

	Penetration Rate (%)		Herfindahl Indices	
	1997	2000	*1997*	2000
Belgium	*9.6*	21.5^		
Germany	*10.1*	58.7	*0.346^*	0.341
Denmark	*27.3*	62.9	*0.509*	0.296
Greece	*8.9*	56.5	*0.514*	0.338
Spain	*10.8*	62.0	*0.635**	0.418
Finland	*42.0*	72.0	*0.562*	0.487
France	*19**	49		0.380
Ireland	*9.5*	73		0.529
Italy	*20.5#*	42.8#^		0.399
Luxembourg	*16.1#*	39.3#^		
The Netherlands	*10.8#*	60	*0.531**	0.343
Austria	*14.3*	49^		0.423^
Portugal	*15.4*	67		
Sweden	*35.8#*	71	*0.279*	0.376
United Kingdom	*20.4*	28.9^	*0.299**	0.260
Norway	*38*	74.8		0.503

Source: the various national regulatory authorities, except where marked with # - these data are from OECD Report 1999 Cellular pricing structures and trends

**1998 data, ^1999 data. The Herfindahl index is calculated using proportion of total subscribers to determine market share.*

There are concerns however that not all aspects of the mobile telecoms market are competitive. Witness the recent raids on German and UK mobile operators as part of an ongoing review on

[2] The index ranges from one, which indicates a monopoly, to zero. A lower value indicates a less concentrated market. If there are n firms each with an identical market share, the Herfindahl index value will be 1/n. If there are four firms in the industry, the Herfindahl index could not be lower than 0.25.

roaming prices by the EC's DG Competition. There are also concerns that the charges for terminating calls to mobiles are too high[3] and that mobile virtual network operators (MVNOs) and independent access providers are being charged too much. These concerns raise questions about the amount of competition in the wholesale market for mobile telephony.[4]

Under regulatory framework provided by EU law and Commission guidelines, National Regulatory Authorities (NRAs) in Member States have acted with a range of measures to regulate competition in the wholesale and retail markets. A number of mobile operators in Member States for example are obliged to offer cost-oriented mobile termination tariffs. Whilst the cost structure of an incumbent's fixed network has been subject to detailed study, the cost structure of a mobile network, and its relationship to prices, is less well understood.

The central purpose of this project was to establish the nature of the cost structures of mobile networks, to analyse the relationships between the prices and costs of mobile services, and to make suggestions for possible Commission recommendations and/or guidelines on costing and pricing in mobile networks. The role of the project was not to assess the competitiveness, or otherwise, of the mobile telecommunications industry in any particular member state or across the EU as a whole. Nor was it to build a cost model to generate estimates of the costs of different services.

Instead the report seeks to provide a clear understanding of the cost structure of 2G mobile networks and their relationship to prices. The focus is on the conceptual issues and on their implications for regulators. Cost data are as sensitive in this industry as in other sectors, if not more so, and no mobile operator was prepared to provide data to us. We have used the limited publicly available data to illustrate key points, and have explained the kinds of data that would be needed for particular objectives or decisions.

1.2 Purpose and Structure of this Report

The rest of the report has the following structure.

Section 2 serves as an introduction and summary to regulatory actions and initiatives relevant to mobile telephony by NRAs in Member States within the framework provided by the EU. There is a discussion of the implications of designation of Significant Market Power. It also addresses other measures that NRAs have adopted to promote competition in the wholesale and retail mobile market. Section 2 is supplemented by a more detailed report of the mobile market and regulatory initiatives in each Member State contained in the *Annex*.

Section 3 identifies the cost elements and cost drivers for GSM networks. What are the factors that influence the total cost of a network? The three standard cost drivers associated with mobile networks are subscribers, coverage and traffic. The section discusses how these cost drivers

3 For example, MCI Worldcom have produced cost estimates suggesting the costs of call termination are much lower than the price. The numbers we have seen are too aggregated to permit detailed analysis.

4 Interestingly, some economists have developed models suggesting that in an unregulated setting, termination charges will be higher the less concentrated the mobile network market is since firms with small market shares cease to consider the demand-reducing effects higher termination rates will have on their business. See Joshua S Gans and Stephen P King (2000) "Mobile network competition, customer ignorance and fixed-to-mobile call prices", mimeo, University of Melbourne.

affect different costs in the network. It also identifies other potential cost drivers. *Appendix A2* to this report introduces in detail the operations provided by a GSM network and the network elements involved.

In *Section 4* the report addresses the question of how to define different markets for mobile services. The distinction between retail and wholesale markets is one that is followed throughout the report. This Section also defines call origination, call termination and roaming services.

The next three Sections focus on the domestic network.

In *Section 5* the possibility of costing the network at a more disaggregated level than the whole network is considered, i.e. what approach can be used to cost individual services? It looks at approaches to allocating costs to services, including a discussion of how to the increments might be defined if seeking to measure incremental costs.

Section 6 discusses how to cost mobile networks given the nature of competition between mobile network operators. It discusses how the relationship between the network's dimensioning and the quality of service might be considered when attempting to cost a given service. The section also discusses how to treat investments, looking at how to calculate annualisation charges and also the effect of future plans on the current costs.

Prices, and how they might relate to costs, are discussed in *Section 7.* There is a discussion of how pricing issues relate to the development of a regulatory framework.

Section 8 looks at roaming. Although the previous sections focus on a domestic network, many of the principles apply equally to roaming. Moreover, the section shows why, from the supply side there is not necessarily a great distinction between roaming services and other services that MNOs provide.

Finally, *Section 9* briefly reviews some of the main implications our report has for regulatory strategies.

2 REGULATORY ACTIONS AND INITIATIVES

Summary

- 24 MNOs have been designated as having SMP in the mobile market. Of these, 12 MNOs in 8 Member States have also been notified with SMP in the market for interconnection.
- Obligations vary with respect to interconnection and site sharing in the Member States.
- Price controls on mobile termination charges apply in 8 Member States.

2.1 Overview

The EC provides the framework for regulatory action and initiatives in Member States. Until the 1999 Review the main body of mobile sector specific legislation relevant to this study could be found under the Mobile Directive (96/2/EC) and the Interconnection Directive (97/33/EC). Other sources of reference include the R&TTE Directive (99/5/EC), the Licensing Directive (97/13/EC) and the UMTS Decision (128/1999/EC).

The 1999 Communications Review for a new Regulatory Framework for Electronic Communications and Services resulted in six proposed legal measures. The Lisbon European Council of March 2000 called on Member States to make the utmost effort to adopt these measures as soon as possible in the course of 2001. The six proposals are to be adopted jointly by the European Parliament and the Council under Article 251 of the EC Treaty. The six proposals are as follows:

1) A common regulatory framework for the electronic communications networks and services Com (2000) 394;

2) Universal service and user's rights relating to electronic communications networks and services Com (2000) 392;

3) Access to, and interconnection of, electronic communications networks and associated facilities Com (2000) 384;

4) The processing of personal data and the protection of privacy in the electronic communications sector Com (2000) 385;

5) The authorisation of electronic communications networks and services Com (2000) 386; and

6) A regulatory framework for radio spectrum policy in the European Community Com (2000) 407.

Of these six proposals, the first three are of particular relevance to this study into the costs of mobile networks. Under (1) (the framework directive) there is a new definition of SMP based on the competition law of dominance. This directive also covers the granting of rights of way and co-location. Under (2) (the universal service and user's rights directive) number portability is to be introduced in Member States, and the right of NRAs to impose proportionate obligations on MNOs to meet legitimate public policy considerations is reaffirmed. Under (3) (on access and interconnection) there is much continuity with former interconnection obligations and the right of NRAs to impose proportionate access obligations on MNOs to meet legitimate public policy considerations is further harmonised and reaffirmed. Specifically, the primary interconnectivity rule and mention of the obligations of transparency, non-discrimination, accounting separation, obligations of access to specific network facilities and cost orientation as available regulatory tools for NRAs.

Table 2.1 presents an overview of the regulatory actions and initiatives undertaken by NRAs in Member States under the current regulatory framework. As the table shows the actions of some Member States have been influenced by the new EU proposals even though the bulk of the regulatory action covered in the table will have been implemented under the existing EC legislation summarised in the first paragraph of this overview section.

The table provides information on the number of MNOs, and notified MNOs by Member State. SMP in the Mobile market and SMP for Interconnection are MNO designations specified in the Interconnection Directive (97/33/EC) based on share of revenue. There is a more detailed discussion of SMP and the trigger obligations that apply in Member States in Section 2.2.

Eight NRAs have imposed some form of price regulation on mobile termination. These are listed in the table and also discussed in more detail in Section 2.3.

Regulators have taken action to promote competition in the wholesale and retail markets. As the table indicates the requirements placed on MNOs to provide wholesale access and co-location or site sharing are not the same in different Member States. To promote competition in the retail market many member states have taken steps to implement number portability in accordance with the proposal on universal service and user's rights (Com (2000) 392.) Measures to promote competition are the subject of Section 2.4, while Section 2.5 emphasises the important choice between different possible objectives that regulators face.

This Section is supplemented by the Annex to this report, which presents more detailed information on the mobile market and regulatory initiatives in each Member State. The data has been collected from a number of sources, primarily from public information made available by the NRAs themselves. Another main source is the implementation reports published by the Information Society on an annual basis. The information is yet to be confirmed with the relevant NRAs.

Table 2.1: Summary of Regulatory Actions

Country	No. MNOs	No. 3G licenses	Notified MNOs		Price controls for call termination *		Measures to promote competition in the wholesale market…		and the retail market
		(of which new entrants)	SMP Mobile	SMP Intercon-nection	Fixed-to-mobile	Mobile-to-mobile	Access and national roaming	Site-sharing	Number portability
Belgium	3	3	1	1	RPI – 15% on termination charges from Feb. 2001		Obligations on national roaming inserted into 3G licenses. New entrants may roam for 8 years.	Site-sharing obligations inserted into 3G licenses.	Introduced 2000.
Germany	4	6 (2)	0	0	-	-	All MNOs must offer interconnection and access for service providers, including MVNOs under commercial negotiations. Co-operation including extensive national roaming endorsed for 3G entrants tp provide 50% coverage.		Scheduled for Jan. 2002.
Denmark	4	4	2	0	-	-	All MNOs must provide access for service providers, including MVNOs, and national roaming subject to technical and financial feasibility. 3G entrants entitled to roam on 2G networks and between one another.	All MNOs must offer the lease of infrastructure and sharing of facilities. SMP MNOs must offer at cost-oriented prices.	Introduced 2000.
Greece	3	3	0	0	-	-	All MNOs must offer national roaming.	All MNOs required to offer the sharing of equipment and property.	Scheduled for 2003.
Spain	3	4	2	1	-	-	SMP MNOs offer access at cost-oriented charges. All MNOs offer roaming to 3G entrants		Not yet available
Finland	4	4	2	2	Sonera has been obliged in April 2001 to lower its interconnection charges to a cost-oriented level.		All MNOs offer access, including for MVNOs. 3G MNOs have right to roam on GSM networks outside their coverage area for 8 years.	All MNOs required to lease capacity and facilities subject to feasibility.	Not yet available

Country	No. MNOs	No. 3G licenses	Notified MNOs		Price controls for call termination *		Measures to promote competition in the wholesale market…		and the retail market
		(of which new entrants)	SMP Mobile	SMP Intercon-nection	Fixed-to-mobile	Mobile-to-mobile	Access and national roaming	Site-sharing	Number portability
France	3	2	2	2	In Oct. 2000 Orange was ordered by ART to lower its call termination charges by 20%. In Nov. 2001 ART ordered Orange and SFR to price-cap call termination to max. 15 Euro-cent/min by 2004.			There are conditions for infrastructure sharing in sparsely populated areas.	Implemented January 2001.
Ireland	3	4 will be awarded by beauty contest at the end of 2001.	2	2	-	-	MNOs must meet requests for access, including MVNOs under commercial negotiations	Subject to commercial negotiation.	Scheduled for 2002.
Italy	4	5 (2)	2	2	18.65 Euro-cents/min SMP only	-	SMP MNOs must offer national roaming for new 3G entrants on cost oriented bases.	SMP MNOs must meet all requests on cost-oriented terms of SMP MNOs.	Introduced June 2001
Luxembourg	2		2	0	-	-	No obligations for access or roaming	.	Not yet available.
Netherlands	5	5	2	0	-	-	SMP MNOs must meet all requests for access, including MVNOs. Roaming subject to commercial negotiation.	Required of all MNOs, but changes to legislation expected in order to strengthen NRA's capacity for enforcement.	Implemented February 2000
Austria	4	6 (2)	2	1	3 operators have been obliged to provide termination at 13.8 Euro-cents/min.		MNO with SMP in interconnection must provide interconnection, including indirect access, on cost-oriented terms.		Not yet available.
Portugal	3	4	2	0	Price cap All MNOs.	-	All MNOs subject to general obligation to lease excess capacity on cost-oriented terms.	All MNOs obliged to share facilities in some circumstances.	Scheduled 2002.
Sweden	3	4 (2)	1	1	0.98 SEK/min (approx. 10 Euro-cents/min) SMP only. All other MNOs allowed additional 10%		All MNOs required to offer excess capacity to service providers on fair market terms. MNOs operating more than 5 years must grant national roaming on fair terms.	Subject to commercial negotiation.	Scheduled Sep. 2001.

Country	No. MNOs	No. 3G licenses	Notified MNOs		Price controls for call termination *		Measures to promote competition in the wholesale market…		and the retail market
		(of which new entrants)	SMP Mobile	SMP Intercon-nection	Fixed-to-mobile	Mobile-to-mobile	Access and national roaming	Site-sharing	Number portability
United Kingdom	4	5 (1)	2	0	11.7 £/min (approx. 19 Euro-cents/min) (1999) RPI-9% SMP only Proposed RPI-12% (2002-06) all operators**	-	SMP MNOs obliged to supply wholesale airtime, and national roaming under commercial negotiation.	Subject to commercial negotiation.	Implemented 2000.

Where an MNO has been designated with SMP for Interconnection, it is also included in the column for SMP Mobile. This follows from the definition of SMP based on share of revenues in specified markets.

The information contained in this table is derived from a number of sources. It has not yet been checked with the relevant NRAs.

** Charge controls are quoted in the relevant domestic currency. Following confirmation with NRAs they will also be presented in euros.*

*** Current proposals are for a change to RPI – 12 per cent for fixed-to-mobile to be introduced in March 2002 when the current control expires, and for this to apply to all four operators.*

2.2 Notified Operators

Until now, SMP has been imposed upon MNOs in Member States deemed to have market share revenues of greater than 25%, under the Interconnection Directive (97/33/EC). Designation of SMP in the mobile market is based in an assessment of mobile interconnection revenues only, whilst SMP for interconnection is based on an assessment of total interconnection revenues from the fixed, mobile and leased line markets considered all together.[5]

Designation with Significant Market Power is the main mechanism by which NRAs become directly interested in the relationship between costs and prices of a particular MNO. Following the 1999 Communications Review, an MNO will be designated with SMP if independently or jointly with others, it enjoys a position of economic strength affording the power to behave to an appreciable extent independently from competitors, customer's and ultimately consumers. Indeed, the revised definition of SMP also covers the leverage of dominant position on an associated market. The proposed directive on access and interconnection (Com (2000) 384) reaffirms previous EC policy that NRAs may impose price controls, including obligations for cost orientation and obligations for cost accounting systems, in situations where market analysis indicates existing and durable market failure.

Table 2.2 lists the MNOs designated with SMP in the Mobile market and SMP for Interconnection in each Member State.

Table 2.2: Notified Operators in the Member States

Country	**SMP in the Mobile market** *(transparency and non-discrimination obligations)*	**SMP for Interconnection** *(cost-orientation obligation)*
Belgium	Proximus	Proximus
Denmark	Tele Denmark, Sonofon	
Spain	Telefonica, Airtel	Telefonica
Finland	Sonera, Radiolinka	Sonera, Radiolinka
France	Orange, SFR	Orange, SFR
Ireland	Eircell, Esat Digifon	Eircell, Esat Digifon
Italy	TIM, Omnitel	TIM, Omnitel
Luxembourg	LUXGSM, Tango	
Netherlands	KPN, Libertel	
Austria	Mobilkom, max.mobile	Mobilkom
Portugal	TMN, Telecel	
Sweden	Telia	Telia
United Kingdom	Vodafone, BT Cellnet	

Source: DG Information Society

[5] See "Determination of organisations with significant market power (SMP) for implementation of ONP Directives", DGINFSO, 1.3.99 at http://www.europa.eu.int/ISPO/infosoc/telecompolicy/en/comm-en.htm#misc.

The designation of SMP for an MNO in either the mobile market or the national market for Interconnection will result in a set of obligations. The designation of SMP in the national market for Interconnection obliges the MNO to observe *cost orientation* for call origination and termination (in addition to other obligations imposed by the NRA.) It is the responsibility of the NRA to monitor and enforce compliance with this ruling. Table 2.3 includes a summary of what initiatives NRAs have undertaken following the designation of SMP for Interconnection.

Table 2.3: Actions Following Designation of SMP for Interconnection

MNOs designated with SMP for Interconnection	Relevant action by regulator
Spain (Telefonica)	SMP MNO (and Airtel) required to unbundle interconnection and mobile services, and undertake cost accounting.
Finland (Sonera and Radiolinja)	Investigated costs for termination in both networks with onus on MNOs to show how they have computed their costs.
	Charge reductions for call termination imposed on Sonera only.
France (Orange and SFR)	Set of guidelines published in June 2001 but assessment of costs and their relationship to termination charges not yet undertaken.
	Orange and SFR ordered to reduce call termination charges to 15 Euro-cent/min by 2004.
	(For all MNOs regardless of SMP designation, ART have fixed call termination rates for incoming international calls. Raising them to FRF 1.26 for 2001.)
Ireland (Eircell)	There has been no regulation of call termination rates. In 2000 the ODTR announced it was undertaking a study of Interconnection rates.
Italy (TIM and Omnitel)	SMP MNO's call termination rates are capped at average charge of 360 Lit/min based on European best practice but these do not presently bite. The MNOs are allowed to differentiate between peak and off-peak termination charges.
Austria (Mobilkom)	SMP MNO has price cap based on NRA's assessment of cost. (Secondary legislation stipulates specifically LRIC basis for cost-orientation of charges.)
	(Other MNOs have been asked to reduce charges in line with SMP's price cap. This is being challenged in the courts.)
Sweden (Telia)	NRA has imposed price reductions for call termination three times on SMP MNO. Price cap is 0.98 kroner/min. as of June 2001.
	(Reasonable termination charges for other MNOs is considered to be the SMP MNO's price cap + 10%.)

Source: NRAs in member states, Info Soc. implementation reports.

2.3 Price Controls

Overall, the market for call termination is directly regulated in eight Member States. The obligation of cost orientation for an MNO with SMP for Interconnection is most directly enforced by NRAs in the market for call termination. This is done by the imposition of a price cap. However, in some member states price controls are not restricted to MNOs with SMP for Interconnection.

Whilst some NRA regulate both fixed-to-mobile termination and mobile-to-mobile termination rates with one price cap, others apply price control only to fixed-to-mobile termination, allowing MNOs to negotiate mobile-to-mobile interconnection agreements between themselves. Both Italy and the UK (discussed in greater detail below) belong to this second category.

In France, the NRA has directly regulated only the fixed-to-mobile termination charge, but has imposed the price cap on all MNOs. International re-routing of national fixed-to-mobile calls (known as tromboning) was becoming widespread due to cheaper rates for terminating international calls on mobile networks. Both national and international termination charges are now required to be at the same level.

In Sweden, the NRA indicated in 1999 that it would not take action against tromboning. The Swedish NRA however, imposes a price cap for mobile-to-mobile as well as fixed-to-mobile termination on the SMP MNO. PTS has reduced the charges of the SMP operator three times since commencing the review of its termination charges in 1999. The termination charge of non-SMP MNOs need not be cost-oriented, however the PTS has indicated that it considers reasonable compensation to be within an additional 10% of the cost-oriented price faced by the SMP MNO.

2.3.1 Establishing the level of a price cap on call termination: Italy and the UK

In Italy, two operators were designated with SMP for Interconnection in Sep. 1999. A price cap was applied to fixed-to-mobile termination charges while no specific regulation was imposed on mobile-to-mobile interconnection charges. The current regulatory regime concerning the calculation of interconnection charges is based on the FDC accounting methodology. The OECD Report on Regulatory Reform in Italy, 2001 indicated that in the fixed network, this tends to over compensate the incumbent by subsidising inefficient historical costs incurred by the incumbent. The interconnection procedure allows AGCOM to determine interconnection charges either by accepting the incumbent's proposal based on the FDC model or to impose EU best practice if the proposed interconnection charge is above EU best practice.

AGCOM has set up a charge cap on fixed-to-mobile termination for the two SMP operators of 360 Lit/min, this is set out in Delibra 338/99. The MNOs are allowed to differentiate interconnection charges based on the retail charging mechanism, so charges differ between peak (430 LL/min) and off-peak (190 LL/min.) The imposition of the price cap resulted in an average reduction in charges of 29%. The level of the price cap was based on an assessment of an international best practice as identified by an external auditor into an investigation conducted by the DG Competition of the European Commission. Also considered was a 1998 KPMG Report for the European Commission suggesting possible differences in termination costs which could be reflected in the termination charge. The price cap that was finally imposed is higher than the one recommended by the competition authority.

In its Recommendation 98/195/EC on interconnection pricing and cost-accounting, issued Jan. 1998, the EC recommended the use of LRIC as the basis for determining interconnection prices to promote efficiency and effective competition.

Another decision by AGCOM, Delibra 340/00 obliges the SMP MNOs to prepare a regulatory accounting system based on FAC for the year 1999, as an intermediate step for the final adoption of LRIC in future years.

Oftel imposed a price cap for fixed-to-mobile call termination rates for the two SMP MNOs in March 1999. The price cap was to be reduced by RPI-9% on an annual basis. The existing controls on Vodafone's and BTCellnet's termination charges are based on a FAC historic cost basis. However, the price-cap is due to come up for review next year. Oftel has conducted an analysis of competition in the mobile market in order to determine whether regulation is necessary for fixed-to-mobile and/ or mobile-to-mobile termination charges. As in Italy, presently mobile-to-mobile termination charges are not regulated. Oftel started a work programme with the industry to establish the LRIC of termination on mobile networks. It has developed a 'bottom-up' economic-engineering model building on earlier work conducted in 1998.[6] Current proposals are to change the fixed-to-mobile price control to RPI-12% from March 2002, and for this to apply to all four operators.

2.4 Measures to Promote Competition

2.4.1 The wholesale market

Not all regulatory initiatives commit the NRA to direct consideration of the relationship between prices and costs. In general, MNOs designated with SMP in the Mobile market are at a minimum obliged to be non-discriminatory with respect to interconnection, so they must offer similar terms to all parties. Within the legislative framework provided by the EC, the NRA can enforce obligations on MNOs to encourage competition regardless of their overall market influence and whether or not they are notified of SMP.

The primary interconnectivity rule implies that all undertakings (in the mobile market) have a right to negotiate terms of interconnection with other undertakings. The definition of interconnection refers to interconnection between two undertakings such that the users of one undertaking may communicate with or access services provided by another undertaking.

Another issue for wholesale competition is co-location of infrastructure. The regulatory framework indicates that site-sharing should be encouraged and subject to commercial negotiation. Collocation of infrastructure may have a significant impact on the cost of rolling out new networks, and the cost of running a network if conditions in a country make it difficult to place and use sites efficiently. In some Member States permits can be refused for environmental and health reasons. In some cases the fees levied for infrastructure development and the taxes levied by municipalities are not harmonised at national level.

NRAs may undertake regulatory initiatives to promote competition by ensuring access on specific terms for new entrants, service providers and MVNOs. Beyond the primary interconnectivity requirement and the condition that collocation should be subject to commercial agreement between MNOs, NRAs across Member States do not have the same additional policies with

---

6 A version of the LRIC model from this OFTEL research was made public on 26 September 2001.

respect to access for MVNOs, national roaming rights for 2G and 3G MNOs, and site-sharing arrangements between MNOs.

Table 2.4 illustrates some of the different regulatory initiatives adopted by NRAs with respect to the wholesale market, but includes only a selection of Member States.

Table 2.4: Regulatory Initiatives on Wholesale Access in Selected Member States

Country	Regulatory Initiatives
Belgium	Obligation for roaming with new entrants for 8 years after they enter the market contained in the conditions for the 3G licenses.
Denmark	SMP MNOs must meet demands for national roaming, and access for MVNOs.
Finland	Obligation to provide interconnection applies to access for MVNOs (though there are no MVNOs operating) and new 3G entrants (for 8 years).
	One MNO without complete national coverage, was explicitly denied the right to negotiate terms for roaming on other MNO networks by the NRA, however all parties have since come to commercial agreement.
Ireland	Roaming for MVNOs is included in the definition of interconnection (though there are no MVNOs in the market).
Italy	Interconnection obligation includes obligation to provide national roaming to new 2G and 3G entrant MNOs.
Sweden	Secondary legislation obliges all MNOs to grant excess capacity to service providers. MNOs who have been operating more than five years are obliged to grant national roaming.

Source: NRAs , Info Soc. implementation reports 2000.

2.4.2 Regulation of the retail market

Handset subsidies have been a significant feature of market development in the retail market of some member states. In the Netherlands the NRA has indicated that competition is being expressed in discounts on handsets rather than in charges for call minutes. In Italy there is no handset subsidisation and this is one reason put forward for comparatively low tariffs in Italy. Handset subsidies are banned in some Member States, for example in Finland.

The new proposed directive on universal service and user's rights indicates that number portability should be extended to mobile telephony. Many countries have taken up this initiative and others have specified a date in the future.

Once number portability is fully introduced, it is more difficult for calling parties to keep track of the rates they are paying to make a call to a mobile, particularly if MNOs charge different termination rates. In Denmark the NRA has indicated that it will enforce measures requiring callers to mobile networks to be informed of their tariffs.

2.5 Identifying Objectives

The recent history of economic regulation in Europe has been predominantly in the context of the regulation of previously state-owned firms, often with significant natural monopoly characteristics. In particular, focus has been on appropriate approaches to price or revenue regulation.

Economic theory suggests that if the regulator has perfect information on the firm and its customers then designing an optimal regulatory regime would be straightforward. But regulators never have such information. Generally the firm has better information about its own costs, its ability to reduce those costs and the preferences of its customers. If the prices a firm is allowed to charge are in some way regulated, there is often a trade-off between the objectives of productive and allocative efficiency.

The goal of realising *productive efficiency* would be concerned with ensuring that the costs incurred to produce a given output are minimised. *Allocative efficiency* is realised if it is impossible to reallocate resources to make one better person better off whilst making no-one else worse off. Any consumer who values a service more than the additional costs that would be incurred serving the consumer should be served. An implication of this is that if prices are above marginal cost for all customers, then allocative efficiency has not been realised.

The trade-off between productive and allocative efficiency as defined can be thought of as potential conflict between incentivising the firm to innovate and approach the efficiency frontier,[7] while at the same time ensuring that prices are reflective of costs incurred by the firm in the provision of its services. Furthermore, price regulation has been affected by equity considerations;[8] concerns over the distribution of income between different groups within the economy are often at odds with attaining allocative efficiency.

A regulator therefore needs to identify priorities. Is it to ensure productive efficiency, or allocative efficiency or to realise certain welfare goals (a "fair" outcome however defined)? It may be that a preliminary analysis suggests that some of these goals can be realised without intervention. In Section 6.1 of this report, we argue that if there is effective competition between mobile operators in the provision of many services, then competitive pressures may suffice in ensuring productive efficiency.

It is also important for regulators to recognise that similar goals in mobile and fixed telephony may not necessarily require the same actions. There are important differences between the two. Typically there are competing networks in mobile telephony, rather than a single dominant firm. And, to pre-empt later sections of this report, the cost structure of mobile networks are significantly different; in particular, the coverage mobile networks provide individual subscribers complicates the analysis of costs. The next sections of this report will explain those differences and show their significance.

---

7 This may be especially important for some previously state-owned firms.
8 Perhaps implicitly rather than explicitly.

3 COST ELEMENTS AND DRIVERS OF MOBILE NETWORKS

Summary

- Mobile networks contain a diverse range of network elements, including handsets, base stations and spectrum, as well as switches and transmission links.
- The three main cost drivers are subscribers, traffic and coverage. It is not always possible to identify a single unique cost driver for each network element.
- Controlling for differences in cost drivers, factors that may cause network costs to differ across countries include differences in terrain, in particular mountains, the amount of spectrum available, and the restrictions on obtaining planning permission.

3.1 Costing Using Published Accounts

As outlined in the previous section, the costs that might want to be calculated will depend on the objective of the regulator. One possibility is that the regulator just wants to know the total costs of an MNO. In this case, a crude first pass would be to look at the costs reported in the accounts. The accounts would not provide accurate information on the economic costs of a network operator because, for example, the approach taken to depreciation may differ to the economic depreciation. Regulators could, and do, overcome such problems by setting guidelines that have to be followed when producing regulatory accounts.

Another possibility is that the regulator might want to calculate the wholesale cost of providing a minute of traffic on the network. An obvious initial way to undertake this exercise would be to look at the accounts of an operator and first identify which of the reported costs are retail and which are wholesale. Having identified the wholesale costs, these could be divided by the total volume of traffic travelling over the network.

Identifying the split between retail and wholesale costs is an exercise that would have to be undertaken for each operator separately, since the ratios of retail to wholesale costs may differ significantly between operators.[9] For those operators that have to produce regulated accounts, provided sensible guidelines govern the production of regulatory accounts, it would be possible to ascertain wholesale costs from these accounts. If the regulator has not imposed the obligation on an operator to produce regulated accounts, then an attempt to ascertain the split could be made by looking at published accounts. Since the data in most financial statements in the public domain will be very aggregated, assumptions will be necessary about how to allocate the costs between retail and wholesale.

[9] Consider two networks that carry the same traffic. If one MNO sells considerably more minutes to MVNOs and airtime resellers than the other, then it is likely to incur fewer retail costs.

At the Commission's request, in the absence of detailed accounting information available from mobile operators, crude cost data estimations constructed in this way are shown in Appendix A1 for several MNOs. The Appendix shows that wholesale and retail costs cannot be distinguished in the companies' financial information, which was provided for quite different purposes. The financial information may include the costs relating to the provision of new services, including 3G, and the depreciation profiles may not be appropriate for regulatory purposes.

There are more fundamental reasons why such calculations will not provide useful cost information for regulation of the price of any particular service. This paper shows that there are significant joint costs in mobile networks. The differences between incremental and stand-alone cost may be large, which may explain why prices can differ substantially between services and across countries (even for similarly sized networks). The real risk of distorting competition and network development by ill-judged intervention makes it important that any regulation of prices in the mobile sector is based on a close analysis of the cost structure of mobile networks and their relationships. This paper seeks to provide the basis for such analysis.

3.2 GSM Network Architecture

As a first step to analysing the cost structure, it is necessary to understand the main elements making up a GSM network and identify the main cost drivers.

The first generation of mobile systems were designed primarily for the transmission of speech signals, although they can also transmit data at relatively low bit rates. They are usually referred to as analogue systems as the speech signals are not digitised prior to transmission by the radio transmitter. These systems include the Nordic Mobile Telephone (NMT), the American Advanced Mobile Phone Service (AMPS) and the British Extended Total Access Communication System (ETACS). Analogue systems are being gradually phased out in favour of digital systems.

The second-generation cellular mobile radio systems are digital, and offer higher speech quality, increased capacity and security, and international roaming — mobile stations can connect and use compatible networks abroad. Examples of second-generation systems are GSM-900 and DCS-1800 in Europe (referred to together in this report as GSM), and Digital AMPS (D-AMPS, IS-54) and PCS-1900 (IS-136, an upgraded specification of IS-54) in the USA.

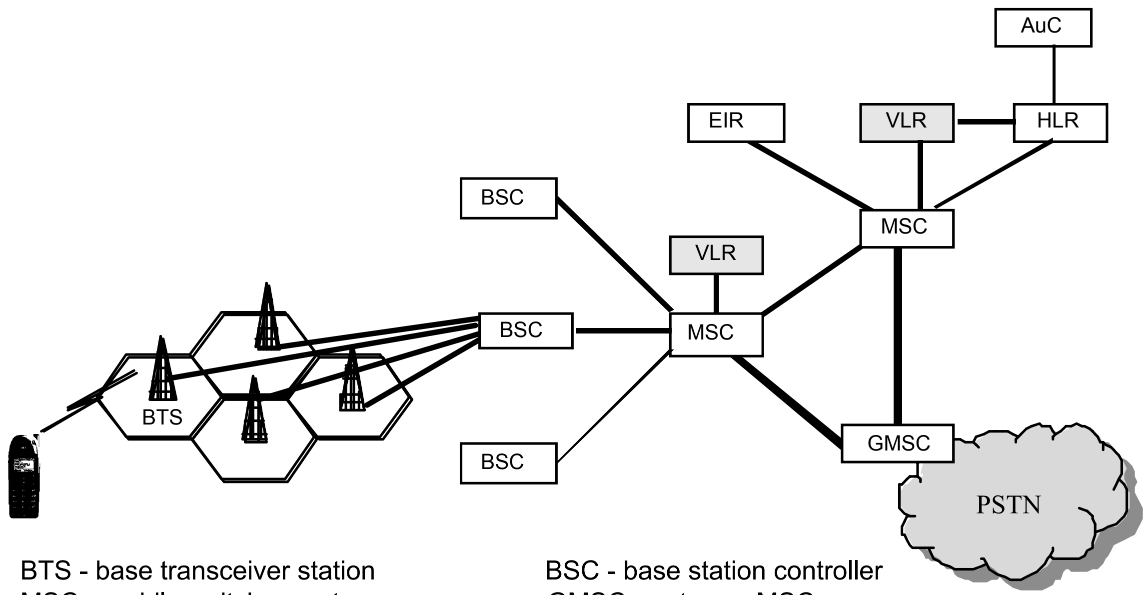
This section provides a brief overview of the workings of a generic Global System for Mobile Communication (GSM) network, describing the main components that make up a network. Some operator's GSM networks differ in significant ways to that outlined below. But the principles are broadly the same. This is also true for 2.5G and 3G networks.

3.2.1 GSM Network Components

The GSM cellular network consists of Mobile Stations (MS) and a fixed support network. The Mobile Station is the handset that people use. Its main function is the voice and data encoding, encryption and transmission (as well as the reverse). The handsets move freely in the network and use radio frequencies to communicate with the fixed support network; they include a SIM card containing details of a subscribers network affiliations. The SIM card belongs to the operator and not to the subscriber, even though the subscriber carries it.

The support network is layered; the degree of stratification and the cell size depends on the user density. Figure 3.1 shows the basic GSM architecture. A simple explanation of what these network components do is given below.

Figure 3.1: Structure of the GSM Mobile System



BTS - base transceiver station
MSC - mobile switching centre
VLR - visitor location register
EIR - equipment identity register
BSC - base station controller
GMSC - gateway MSC
HLR - home location register
AuC - authentication centre

Base Transceiver Station

The BTS provides the radio coverage for a cell in the GSM network. It encodes, encrypts and transmits the voice and data signals to the handset (as well as the reverse); it also broadcasts the control and common control channels. A BTS is connected to a BSC, either directly or via a "daisy chain" through other BTSs.

Base Station Controller

The BSC performs radio resource management, acting as a concentrator and switch ensuring that calls from the MSCs are passed to the correct BTS. The BSC controls the inter-cell handovers for handsets moving between BTSs, the reallocation of frequencies among BTSs, and the power management of BTSs and handsets within the catchment area.

To improve network resilience, there may be dual links between a BSC and its parent MSC.

The BTS and BSC can sometimes be grouped together and called the BSS (Base Station Subsystem).

Mobile Switching Centre

The MSC is in charge of location registration and the dynamic allocation of resources to co-ordinate call set-up. It looks after the handover management, billing, SMS and paging. MSCs are switches that route calls around the network and computers that search for and process information from location registers.

As the number of users increases, full intermeshing of the MSC becomes very costly and complicated. Therefore a second layer, Transit Switching Centres (TSC), has to be added to the hierarchy. These are now present in almost all mobile networks. The transit layer forms the long-haul backbone of a GSM network, carrying calls between MSCs (including to and from the Gateway MSC).

Gateway MSC

The GMSC is an MSC that gives a point of interconnection (PoI) for the mobile network with other networks, e.g. the PSTN or the networks of other mobile operators. Interconnected parties may have two or more interconnection points.

Home Location Register

The HLR is a permanent database holding administrative information relating to all subscribers to the mobile network. There are usually only one or two HLRs per network. For each subscriber, two numbers are jointly stored in the HLR: the Mobile Station international ISDN Number (MSISDN) and the International Mobile Station Identity number (IMSI). The IMSI is the subscriber's unique identification number, contained on his/her SIM card. The MSISDN is the directory number dialled to reach the subscriber which, among other things, identifies the network operator and services the mobile user has access to. In addition to this permanent information, the HLR also contains information regarding the current location of each subscriber; this is generally in the form of the address of the Visitor Location Register at which the subscriber is currently registered.

Visitor Location Register

Each MSC in the network is linked to a corresponding VLR. The VLR temporarily stores subscription information on the users currently located in the cells associated with the MSC to which the VLR is attached. This information is obtained from the HLR and allows the subscriber to make use of services (e.g. outgoing calls) without the need for the HLR to be queried each time. Thus the information which enables subscribers access to the mobile network is delegated from the HLR to the VLRs. Additionally, the VLR holds more precise information (than the HLR) as to the location of each handset within the area covered by the MSC/VLR.

Authentication Centre and Equipment Identity Register

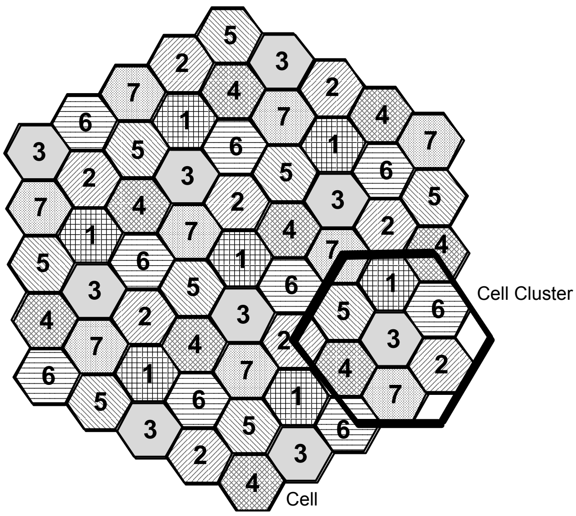
The authentication centre stores the keys that are needed in the encryption algorithms for each subscriber. The equipment identity register keeps information on the handsets currently in use, enabling stolen handsets to be rejected by the network. The AuC and EIR can be implemented as stand-alone nodes or as a combined AuC/EIR node.

3.2.2 Cellular Networks

3.2.2.1 Cell structure and frequency use

Cellular mobile radio makes efficient use of scarce radio spectrum by using low-power (short-range) transmitters that use specific radio channels in an area called a cell and reusing these channels in other non-adjacent cells spaced far enough away so that interference between cell channels is avoided. The minimum set of cells using all the available frequencies is called a clusterset.

Figure 3.2: Frequency Re-use Pattern



Frequencies cannot be re-used in adjacent cells because of problems with co-channel interference but they can be repeated throughout the network following a pre-set frequency re-use pattern (Figure 3.2).

For communication, the base station and the handset use duplex radio frequencies i.e. the communication takes place on a pair of channels separated by 45 MHz. On a GSM900 network the handset transmits to the base station on a frequency from 890-915 MHz (*Uplink*) and receives communication from the base station on 935-960 MHz (*Downlink*). A 200KHz carrier spacing has been chosen. Excluding the two 100KHz edges of the band this gives 124 carriers for *Uplink* and *Downlink*.

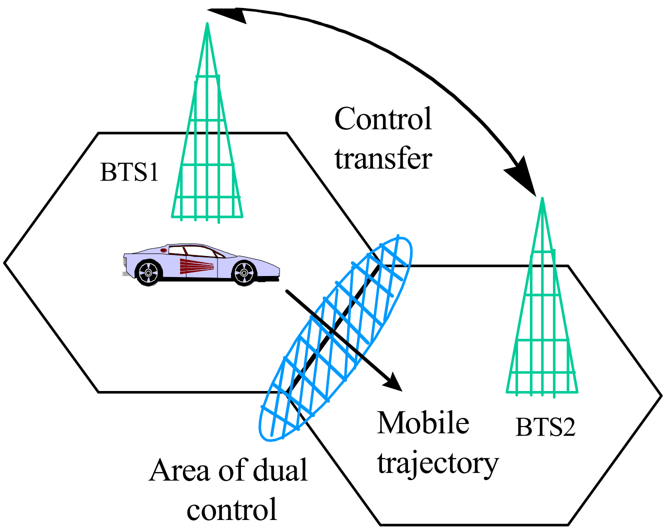
There are two types of channels used by the handsets for communication — traffic channels and control channels. The traffic channels are used for the transmission of encoded speech and user data. For example, when a user is talking, the speech is encoded and transmitted by the handset using the traffic channel. The control channels carry signalling and synchronisation data between the base station and the handset. These channels are used for frequency correction, broadcasting of information to all mobiles in a cell, synchronisation of transmission and to control handovers.

3.2.2.2 Subscriber Mobility

The obvious difference between a mobile network and a fixed network is that subscribers to a mobile network can make and receive calls in different locations. This difference is particularly important when there is an incoming call for a mobile subscriber because the subscriber's location varies. One way around this issue would be for every cell in the network to be paged when there is an incoming call for a particular subscriber. However, this approach would be wasteful in terms of the quantity of bandwidth required in order to send signals across the whole network for every call. Rather, the convention is for cells to be grouped together into *location areas* across which the signal for an incoming call for a particular subscriber is delivered.

Clearly a mechanism is required by the network to keep track of which location area each subscriber is currently in. Signalling is therefore needed to update the network of movements of users from one location area to the next. As a subscribers move between cells in different location areas, they pass into areas managed by different BSCs. This process is called "handover" (see Figure 3.3).

Figure 3.3: Handover Procedure



The smallest possible location area would be an individual cell. However, there is a trade-off: smaller location areas use less bandwidth for signalling incoming calls but more bandwidth for signalling location updates. A typical location area might cover around 100 cells under the control of one MSC.

The capacity of a cell will be particularly important in determining the network's performance in enabling subscriber mobility. The handset can be in one of two distinct modes: it can be in *idle mode*, waiting for calls, or it can be busy with a call or other service in progress (*dedicated mode*). If a user moves to a new cell while the handset is in idle mode, and the new cell is very busy, there is a possibility that the user will be unable to initiate or receive calls. This situation is termed *call blocking*. On the other hand, if a user moves to a busy cell when the handset is in dedicated mode, it is possible that there would be no free channel on which the call can continue. In this case the call would be *dropped*. The number of blocked and dropped calls is a major indicator of the performance of the system.

3.3 The Concept of Coverage

As highlighted above, the fundamental difference between a fixed and a mobile network is that users can make and receive calls in varying locations — they have mobility within the network.

While a fixed network provides a direct one-to-one link connecting each subscriber to the network at a particular location, a mobile network provides coverage. Coverage can be thought of as a service permitting a non-specified subscriber to access the network at a given place. If enough subscribers value the ability to connect to the network from a certain location, then coverage will be provided in that location.[10] As the number of such locations rises, coverage is extended geographically.

Describing coverage as the ability to connect a subscriber's handset to the operator's network still leaves some ambiguity. There are a number of possible definitions that could apply to the concept of providing coverage. Some examples follow.

- A network provides coverage if all customers can connect with the network in that place at all times. Under this definition, the network would require a huge capacity so that even if a great many subscribers are in the same area, perhaps at a major sports event, they can all make or receive a call. Probably no networks would provide coverage under this definition.

- Another view is that coverage relates to the network as it has been designed, since its design optimises the ability of a subscriber to connect with the network, recognising the trade-off between the cost of increasing capacity versus the desirability of being able to connect to the network at any time in any place.

- Coverage is the capability or option to make a single call from any point of the network at a point in time. All additional capacity is due to conveyance.

- Coverage is the site locations necessary for an MNO to construct a network to which people within a defined area could connect to a network. All of the telecom equipment necessary to permit a connection is due to conveyance.

We believe that the definition of coverage in the third bullet is the most suitable for the purposes of costing a network. Increasing coverage means increasing the area within which a connection is possible: the main cost driver is km^2. Having a single major cost driver for coverage helps in terms of identifying cost causality if attempting to cost individual services. In contrast, the first two

---

10 Although in some cases the extent of geographical coverage is stipulated in a mobile operator's licence conditions.

bullets provide definitions of coverage such that the costs of providing coverage would depend both on the area covered and on either the number of subscribers or the number of voice minutes. The final bullet provides a definition of coverage that does not measure the costs of permitting a subscriber access to the network.

Coverage (under the preferred definition) is not necessarily an access service in the sense of fixed-line networks. Nor does the definition of coverage provide definite answers as to how coverage relates to the different services that users of a mobile network actually purchase. These are important points and are returned to in Section 4.

3.4 Cost Drivers in a Mobile Network

One tenet of cost-based pricing is the idea that the person who caused a cost to be incurred should pay for that cost. Hence the desire to establish cost causality for the various costs that a regulated firm incurs. Identifying cost drivers can help in this regard.

A cost driver is a factor that causes variation in the total costs of provision of a product or service. The necessary feature of the driver is that there is a causal link to costs such that if the level of the driver alone changes, this results in a change in the total costs incurred.

There are practical problems that arise when seeking to analyse cost drivers.

- Where the provision of a service is entirely through economies of scope (i.e. all costs are joint costs) no cost driver can exist because costs are not variant to changes in one activity or output alone. This problem is returned to in later sections, when considering how to cost individual services. In this section, the focus is on the cost drivers for the entire network.

- Are cost drivers the final outputs (services) that a firm produces? Or are cost drivers the factors that influence the costs at discreet stages of production?

This section seeks to consider the importance of subscriber numbers, coverage and call traffic as cost drivers. Holding other factors constant, what happens to costs as one of these drivers increases? This exercise illustrates one of the problems with seeking to identify cost drivers. The number of subscribers will, as argued below, affect the number and therefore cost of handsets. It is also likely in practice to affect the number of call minutes, so that it could be argued that the number of subscribers indirectly affects those costs that are deemed to depend on the cost driver of call traffic. But since the objective is often to identify prices that will reflect the costs individuals cause, there is no advantage in arguing that subscribers is the cost driver. The caller who gave rise to a minute of traffic should pay for the costs of that minute, rather than averaging the costs across all subscribers.

For mobile networks, a problem arises because of coverage. As the area covered increases, the costs of the network operator increase. So coverage area qualifies to be called a cost driver. Yet knowing how sensitive network costs are to the area covered does not help with seeking to have prices set according to cost causality. It is not possible to identify a single individual who is responsible for the network covering a given area. And nor is it possible to trace back to argue that the volume of coverage is in turn dependent on another cost driver which can be associated with the actions of a single individual.

Table 3.1: Cost Drivers and Cost Causality

Cost Category	Cost Driver	Cost Caused By	Implication for cost-based pricing
SIM Cards	*Number of subscribers*	Individuals subscribing	Bill each subscriber for the cost of a SIM card
Switches*	*Number of call minutes*	Individuals making calls	Include the total cost of switches divided by total number of minutes in a per minute charge paid by people making calls.
Racks*	Number of switches, which in turn depends on *number of call minutes*	Individuals making calls	Include the total cost of racks divided by total number of minutes in a per minute charge paid by people making calls.
Rural sites	*Area covered*	?	?

* *As discussed below, in practice the cost drivers for many elements in a mobile network could be either number of call minutes, call attempts, or coverage area.*

The focus is on identifying how the network costs will change as a cost driver changes holding all else constant. How much would it cost to carry an extra million call minutes whilst the same number of subscribers in the same coverage area with the same quality of service. In practice, an MNO must decide between incurring the costs associated with upgrading the network or the potential opportunity costs associated with permitting degradation in the quality of service the network provides (i.e. costs resulting from a reduction in service demand and consequently revenue).

3.4.1 Subscribers

Every subscriber needs a handset and associated SIM card. Consequently these are customer-driven costs, not coverage or traffic driven. The costs of the handsets are typically borne by customers, although in many countries operators offer handset subsidies. The cost of such subsidies depends on the number of subscribers, not the number of call minutes.

More (switched-on) subscribers mean that there will be more signalling associated with location updating. So the number of subscribers, as a driver of the costs of location updates, could be an indirect cost driver for all these network elements, although one might argue that location updating only takes place because it is the most efficient way for terminating calls so call termination is the cost driver. However, the volume of traffic caused by location updating is likely to be small and not a major cost driver for most of the elements involved.

But the number of subscribers will act as a more important cost driver for the HLRs and VLR since extra subscribers will require the location registers to manage larger databases and have greater processing power. Similarly, the costs of the AuC and EIR are also likely to depend on the number of subscribers.

Billing costs will also depend largely on the number of subscribers for pay-as-you-go callers. For pre-pay customers it is less clear whether the costs depend on the number of subscribers or the number of call minutes they make.

There is no need for additional MSCs because of the number of subscribers (directly) since there are no resources dedicated to subscribers (unlike in fixed networks where there are line cards).[11] In a fixed network it is possible that a new switch is required even thought the existing switch has spare processing capacity because of a limit to the number of subscribers that can be linked to a switch.

3.4.2 Coverage

To provide coverage requires the ability for the subscriber's handset to connect with the operator's network for location updating. Some spectrum will need to be available in the geographical area to be covered and a BTS located with a sufficiently powerful transceiver that communication between the BTS and handset over the frequency is possible. There will also need to be backhaul between the BTS and the BSC, and ultimately links to the switching network.

There are two ways that coverage can be increased. The first is to increase the power of the existing transceivers so as to increase the area served by each BTS. The alternative is to add a new BTS, along with the backhaul necessary to link the BTS to a BSC. Adding a BTS to the network is a much greater cost undertaking than increasing the range of an existing BTS. So in practice, a network designed to cover a given area will seek to achieve this using the minimum number of BTSs.

Signal degradation means that in practice a BTS can cover an area with a radius of 30 km or more in a GSM 900 network. The area is smaller in a DCS 1800 network; this is because the radio propagation is poorer at 1800 MHz, i.e. the range is shorter. These numbers will be smaller in mountainous and (possibly) hilly regions. This is because mountains will block the radio air interface between the MS and the BTS. A mountainous region may require many more BTSs as is normally required to cover a given area.

There are also other engineering reasons why the costs of covering mountainous regions are higher, similar to those reasons given for why mountainous regions are expensive in fixed networks. For example, laying duct and fibre is expensive, so microwave will be used more extensively than elsewhere in the network.

In terms of providing coverage, the cost difference between urban and rural areas will relate to the difference in getting a site license. This is likely to be more expensive in urban areas. However, as discussed later, in urban areas the main driver determining the number of BTSs is unlikely to be coverage.

---

11 This assumes that the number of subscribers is not so large that extra switches are necessary for location updating.

A BSC can handle approximately 100 BTSs, so a significant increase in the area to be covered may result in an increase in the number of BSCs required in the network, and also in the costs of transmission between the BSCs and the MSCs. Providing extra coverage will not be a major cost driver in terms of the number of MSCs required in the network. It is the “radio network” for which coverage is a major cost driver; but traffic is also an important cost driver for antennas, BTSs, and BSCs, as discussed below. Since the radio network accounts for the majority of costs in a mobile network, this creates problems if attempting to cost services and apply cost-causality principles.

A crude estimate of coverage costs would first estimate the total land mass covered by the network and divide this by the maximum area that can be covered by a single cell. This would provide information on the number of sites that would be necessary. So for an Austrian GSM-900 operator that covered all of Austria, coverage might require 400 sites.[12] The cost of a site would include the costs of BTSs, antennas, and leases on towers and land. These costs will constitute the majority of coverage costs. There will also be some costs for transmission, BSCs and MSCs, but these will be relatively small.

Therefore, differences in the stand-alone coverage of different networks will largely depend on the land mass covered. They may be twice as large in Austria as in the Netherlands. Some caveats apply. First, an 1800 operator is likely to need more than double the sites for a given land area so their coverage costs will be almost doubled. Second the terrain covered by a network may differ, with the implications this may have for coverage. Networks covering mountainous regions will need more sites than the crude estimate suggested above implies. Third, leases to locate equipment may differ significantly between countries, although these leases will probably account for less than 20 per cent of a site’s cost. And finally, health and safety regulation may either require more sites than suggested above or may increase the costs associated with setting up and maintaining a site.

3.4.3 Traffic

The differences in the way the network is used for call set-up and call duration is largely ignored here when asking how network costs will vary as the volume of traffic increases. The implicit assumption here is that an increase in traffic means the same proportionate increase in both the number of calls and the number of call minutes. The distinction is probably most important when considering the costs of MSCs, for which call attempts is a cost driver, and not just call minutes.

There are three approaches to increasing the capacity of a mobile network. One is to increase the capacity of existing elements in the network, but this strategy is constrained by the amount of available spectrum. When the constraint becomes binding the alternatives are either to increase the number of cells to permit more use of the existing spectrum or to acquire more spectrum.

---

12 83,000 km^2 divided by 210 km^2, the approximate area covered by a GSM 900 rural site in a flat region. This assumes that there are no “shadows” caused by mountains, i.e. for illustrative purposes Austria is assumed in the example to contain no mountains.

3.4.3.1 Increasing capacity of existing elements

At low levels of demand the solution to serving extra traffic may be to augment the existing equipment.

New frequencies could be made available at the cells. This would result in additional costs associated with the BTSs. The ability of a BTS to carry traffic is limited by the capacity of its processor unit. This unit translates messages into formats suitable for MSCs or handsets respectively, amplifies signals broadcast over the air interface and manages handover from one sector to another sector within the same cell site. As traffic volume increases, the processor's capacity is exhausted and quality of service reduces. To maintain a desired quality of service will require increasing the capacity of the BTS. Increasing the capacity of the BTS may feed through and affect the costs of backhaul, BSCs, transmission and MSCs.

The BSCs monitor the BTSs and allocate BTS resources to calls, as well as handling call hand-offs from one cell site to another as a subscriber moves from one cell to another during a call. The BSC is also responsible for voice coding, which permits high-speed transmission of voice in a wireless network. These features depend on the volume of traffic. When a subscriber makes a call, some of the BSC's capacity is unavailable for other calls until the call is disconnected.

Bandwidth on the backhaul links is assigned to calls as needed, rather than being dedicated to a single customer. As traffic volumes increase, the amount of backhaul bandwidth necessary to maintain the desired quality of service increases. This will increase the costs of backhaul, since the rates for higher capacity links will be higher — a 2Mbit/s link is cheaper than an 8Mbit/s link. Hence the costs of backhaul are traffic sensitive, although there are modularities, e.g. a 2Mbit/s link will suffice for a range of traffic volumes.

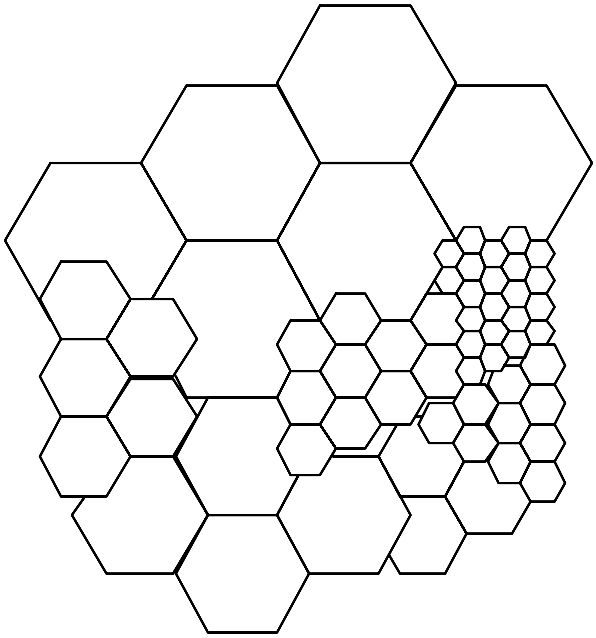
The signalling network is made up of packet switches and signalling links. These have limited capacity to process and transport messages during call set up. Consequently, additional capacity will be necessary to ensure an acceptable proportion of calls are realised in peak periods. Similarly, where advanced services are offered from a mobile intelligent network (IN) known as CAMEL (Customised Applications for Mobile networks Enhanced Logic), this element will require enhancing as traffic levels increase.

The central processor and switching matrix of an MSC, along with the ports connecting the MSC to other parts of the network, are all traffic sensitive. To maintain an acceptable quality of service with regards call set up, routing, and switching requires increasing the capacity of MSCs as the volume of traffic increases. A single switch may not be sufficient in densely populated metropolitan areas to process all the calls.

3.4.3.2 Cell splitting

The number of channels in a cellular network can be increased by the use of smaller cells. An average cell, for example, contains 24 channels, but for some busy areas this is not sufficient. This shortage can be overcome by splitting the cell into smaller cells (shown in Figure 3.4) through limiting the transmission power of the BTS and then re-using frequencies — the BTS's can have a radius between a few hundred metres to 30 km or more.

Figure 3.4: Cell Splitting



Urban areas have more cells per square kilometre than rural ones and the trend towards smaller and smaller coverage area is likely to continue in the future with the development of micro-cellular based systems. It is possible that a single picocell might cover a large administrative building or block of flats. In some cases cells may completely overlap. For example, one cell may cover a town, but a microcell may be installed to cover the town centre to provide additional coverage.

As the cell radius (assuming hexagonal cells) decreases the probabilities for *call blocking* and *call dropping* also decreases. Therefore a reduction of cell size will have a positive effect on the performance of the system. In theory, the splitting of cells and reuse of frequencies can increase the capacity of the cellular system infinitely. However, although cell splitting increases the number of available channels, it also increases the total cost incurred by the network provider.

Increasing the number of cells will result in an increase in the number and therefore costs of BTSs, antennas and supporting towers. More BTS processors and antennas will be necessary. Since the antennas will typically be located in a tower or on a rooftop to ensure adequate signal strength, more towers will be required. The tower's height, the size of the antenna and the rental for the cell site will largely depend on where the cell is located.

Where it is possible to share the costs of the tower with other operators, a forward-looking model should not assign all the costs of a tower to a single operator, even if currently the operator is the only one using the tower. Instead the cost of the tower is what it costs to lease tower space or the costs of the tower less revenues earned by leasing to other operators.

The increase in the number of BTSs will increase the costs of backhaul, since more links will be necessary. Two backhaul links will be more expensive than a single link carrying twice the traffic load.

Another consequence of the additional BTSs that arise if the MNO splits cells is that more BSCs may be necessary. This is a second reason why the cost of BSCs is traffic sensitive.

3.4.3.3 Adding spectrum

In the long run, when all inputs are variable, the engineering design that minimises the cost of the network will involve judgements about the optimal combination of spectrum and cells: less spectrum will require more cells to permit greater frequency reuse; more spectrum means fewer cells since larger cells are possible.

The need for spectrum and cell sites depends in both cases on the volume of traffic. For a given amount of spectrum, only a limited number of calls can simultaneously be made holding cell size constant. As traffic grows, more spectrum is necessary. None of the spectrum is dedicated to a single customer, so it cannot be argued that the costs of spectrum depend directly on the number of subscribers. The costs of spectrum are traffic driven (subject to the caveat that some spectrum will be necessary to provide coverage).

What are the costs of spectrum? If it cannot be traded, or if it comes in large non-divisible bands of frequency, it is possible that even in the long run the cost to the MNO of the spectrum necessary for coverage is the same as the cost of the spectrum necessary for the level of traffic on the network. Of course, this cost may be zero in countries where the spectrum is awarded to operators. This ignores the opportunity cost associated with using the spectrum to provide mobile telephony services rather than another service, but most licenses require that the spectrum be used to provide mobile telephony services. The MNO could not put the spectrum to other more profitable uses even if they existed. But if free spectrum trading is allowed, the opportunity cost of the spectrum used by an operator will be reflected in its market value.

3.5 Implications

The most useful way to think about the cost structure of a mobile network will depend in part on what the purpose of the exercise is.

It is possible to undertake “back of the envelope” calculations. But for most regulatory purposes such exercises will not be sufficiently robust. And often they will not be necessary since the regulator has the power to collect more detailed information and undertake a proper analysis. For example, dividing total costs by total traffic volume is unlikely to produce cost estimates that can be usefully compared with prices, as it would fail to identify the cost of any given service. If the concern is that excess profits should not be earned, it is simpler to compare total revenues and total costs.

The most obvious difference between fixed and mobile networks when thinking about the costs is the existence in mobile networks of a third important cost driver, coverage. In later sections of this report it is frequently coverage that complicate attempts to implement the sorts of rules developed for thinking about cost structures in a fixed setting. There are different ways to think about coverage. We recommend that coverage costs be thought of as those costs that arise in ensuring that a subscriber can make or receive a call from anywhere in the area covered. The debate about how to define coverage may seem philosophical. Regardless of definition, there will be practical problems when attempting to cost a mobile network that, ultimately, relate back to the fact that one of the main cost drivers for a mobile network is geographical area covered.

The costs of many network elements will depend on two cost drivers, the volume of traffic and area covered. This makes it difficult to undertake simple sensitivity analysis. It is not possible to determine how a 10 per cent increase in one of the cost drivers will cause the costs of providing that network element to increase by X per cent, without specifying the starting point, i.e. the volume of traffic and geographic coverage provided before one of these cost drivers was increased by 10 per cent.

It cannot be argued that an individual through his or her actions caused an increase in the geographical coverage of the network. Yet the coverage area is an important cost driver. Attempts to price on the basis of cost causality will have to address this fact.

The costs with providing coverage are different for 900 and 1800 operators (they are higher for the latter because cells can cover smaller areas). This may have implications if seeking to develop regulation that is "technologically neutral". The decisions of a regulator on how a network operator can recover costs due to coverage may affect the ability of operators at these different frequencies to compete.

4 MARKETS FOR MOBILE SERVICES

Summary

- There is more than one way to define the services that a mobile network operator provides. For example, in the calling party pays environment is the caller to a mobile network paying for coverage? Yet to cost a service requires identifying exactly what the service is.
- This report assumes that coverage is not a service that people purchase separately. Instead call origination, call termination, and various data services are each bundled services that include both conveyance and coverage.
- For roaming, the wholesale services of call minutes are assumed to include charges for coverage. The retail services include charges for costs such as billing and advertising.

There are arguably three stages to costing services: identify the services; determine a method for allocating the costs to the different services; calculate the costs of each service.

In many instances the services that the mobile handset owner values are also valued by other parties, and in some instances it is the other party who is charged for the service. For example, call termination charges are levied on the caller in the calling-party-pays environment. This distinction about who pays does not affect the costs of providing the service, but it might have implications if considering how competitive provision of the service is.

For a GSM network operator, identifying meaningful services is not straightforward. Although it is a seemingly academic and philosophical exercise, defining the services is a necessary and unavoidable first step to ascertaining the cost of individual services. If it is deemed important to calculate the costs of individual services, then this exercise will have to be undertaken.

A “service” needs to have the properties that there is an identifiable demand for it and, if the service is to be costed, there need to be separately identifiable costs. There are a variety of services demanded by consumers. In theory, there are geographical and time distinctions that can be made to unbundle marketed services. For example, the service of call termination in a country could instead be thought of as two services — call termination in the capital city and call termination in the rest of the country — bundled together. The costs of terminating calls in the capital city may be very different to the cost of terminating calls elsewhere. At the extreme, it is possible to define a call originated from a given location at a given time as a distinct service from all other calls originated or terminated by the network. The usefulness of this exercise is questionable.

This section discusses the challenges inherent in defining services and proposes a working definition for the rest of this report. It starts with some general thoughts on the nature of the retail and wholesale markets in mobile telephony. Then it moves on to discuss what the services of call termination and call origination include. Much of the focus is on determining who pays for coverage. Section 4.3 discusses the services associated with roaming: what is the roamer buying? Section 4.4 provides guidance on an approach to defining services.

4.1 Nature of Retail and Wholesale Markets

At the network level there are barriers to entry because of the scarcity of spectrum. This may suggest that it is worthwhile making a distinction between the wholesale and retail markets, since only a limited number of firms can provide wholesale services. Although the number of firms in the retail market will depend in part on behaviour in the wholesale market, there are no technical reasons that constrain entry into this sector of the market. In addition, the contracting parties tend to be very different in the two markets, and there are differences in the types of costs incurred.

The retail mobile market is the market for goods and services sold to end-users, whether corporate or private. The end-user makes decisions about which handset to buy, what tariff bundle to sign up to, and the range of services to use. Competition in the retail market can take many forms. Firms may compete in prices or in terms of quality or range of services. The nature of competition may differ across countries. In some, there may be fierce price competition; in others competition to win end-users may focus on offering more varied or higher quality services.

Price competition can take the form of different tariff bundles, with customers choosing the tariff bundle that best serves their needs. In some countries, much of the price competition has been in terms of handset subsidies; in Finland such subsidies are banned. There is also competition in terms of the handset functionality and the range of services that the customer is offered — recent developments such as WAP and GPRS are examples of MNOs seeking to be able to offer customers improved data services. Retail customers will also care about the coverage that a network provides, both in terms of its geographic reach and also the “depth” of coverage, i.e. will the call be blocked or dropped?

It is technically possible for customers to purchase calls and other services from a variety of sellers. They do not have to purchase from the MNO whose network their calls originate on. Clearly there will be some constraints on the ability of resellers to compete on aspects other than price: they have to offer the same geographic coverage as the network from whom they are buying their wholesales services. Regulatory actions will affect the nature of competition in the retail markets. For examples, the introduction of number portability or the banning of handset blocking may influence how competitive the market is.

Regulatory actions in the wholesale market may also feed through to affect the retail market. In the wholesale market, there are a wide variety of possible models for intermediaries. These can range from Mobile Virtual Network Operators (MVNOs) who have their own networks but purchase the radio frequency from MNOs, to airtime resellers who act as distributors for the MNO from whom they have bought traffic minutes or volumes. The MNOs are usually vertically integrated companies, selling in both the wholesale market and the retail market. The termination of incoming calls is also a wholesale mobile service, sold by the MNO to the operator of the customer originating the call. Similarly, the sale of international roaming rights is a wholesale service.

The wholesale markets can differ significantly between countries. Regulatory actions may influence developments. For example, rules on access rights and non discrimination have been applied in different ways in different countries. The restrictions, if any, placed on the tariffs that MNOs can offer in the wholesale market may also influence competition. The form of the wholesale tariff may affect the viability of certain retail pricing plans that might otherwise have been a desirable development in the retail market.

4.1.1 Retail services versus wholesale services

Accepting the argument that a distinction should be made between wholesale and retail mobile services, there remains the task of which activities undertaken by the MNO fall into each category. For many items, the attribution is obvious. Provision of network infrastructure to enable mobile telephony (spectrum, transmitters, switches, backhaul etc.) is clearly a wholesale activity. Likewise, final billing of customers and customer relations is clearly a retail activity.

The distinction is not so clear when it comes to advertising. On the face of it, advertising is a retail activity, undertaken by the retail division of an MNO in order to increase its retail revenue from end-users of the network. An MNO that increases advertising expenditure would be looking for the costs of this activity to be met by more subscribers joining the network, hence bringing in higher retail revenue through subscription fees and call tariffs.

But this viewpoint may be too simplified. Suppose a network has its own retail division and also sells wholesale minutes to an independent reseller of airtime. The reseller undertakes all billing and relations with its customers, as well as advertising of its brand. The retail arm of the MNO also undertakes these activities. If the potential customers of the reseller are unaware of which MNO provides the reseller's wholesale minutes then the division is straightforward: the retail arm of the MNO and the reseller advertise independently and it seems fair to attribute the MNO's customer advertising expenditure solely to its retail division.

However, the situation is different if consumers are aware of which MNO provides the reseller with wholesale minutes. Under this scenario, when the MNO advertises its network it may be expecting to increase sales revenue both directly from its final retail customers and indirectly through wholesale revenue from the reseller. The latter would be significant if the reseller's customers valued the network that their calls were delivered across (e.g. through perceived higher quality) and advertising by the MNO increased this valuation. Although advertising of the MNO's retail tariffs would not achieve this effect (these tariffs are not paid by the reseller's customers), advertising by the MNO of its brand or network quality may do. Thus such advertising by the MNO could be perceived — to an extent — as a wholesale activity because it aims to increase (indirectly) wholesale revenue.

More generally, isolating the services that advertising expenditure is incurred to promote will entail judgements over which reasonable people may disagree. For instance, billboard advertising of the MNO's brand may increase subscriptions, but these extra subscribers may also result in more calls to the network, and therefore more call termination revenues. So were the costs of advertising to promote retail services or the wholesale service of call termination? How the costs of advertising should be attributed to individual services may not simply depend on an assessment of the benefits to the firm.

The Principle of Non-discrimination

The principle of non-discrimination currently applies to operators who have SMP in the mobile market. The principle dictates that a designated operator must charge the same rate for an interconnection service to other operators as it charges itself:

The operator "*shall apply similar conditions in similar circumstances to interconnected organizations providing similar services … as they provide for their own services, or those of their subsidiaries or partners…*"[13]

An implication of the non-discrimination principle is that a network operator cannot charge more for an interconnection service than it does for a final retail service of its own that uses such interconnection as a wholesale input.

The primary interconnection service that a mobile operator sells is call termination. It follows that under non-discrimination, the operator cannot set a retail price for a service involving this interconnection below the wholesale interconnection rate.

It is arguable that on-net calls include the same network operation as call termination. Both require a call to be terminated to one of the network's subscribers. Although an on-net call does not come in from another network, it will be received by an MSC (acting the same way as a GMSC) and the same requirement to find and notify the called subscriber exists as for off-net calls. Therefore, firms following the non-discrimination principle, will not price on-net calls below the wholesale termination rate (for the relevant time of day). Note that on-net calls may allow for more efficient routing than off-net calls since in the former the network knows the location of both parties (e.g. consider the case of both parties being within the area served by the same BTS). But this is not a fundamental difference between the two services; regulations applying to "similar services" and "similar circumstances" would therefore appear to catch on-net and off-net calls.

Regulators wishing to check for non-discrimination can compare on-net call charges with operator's termination rates. If SMP operators do price retail on-net calls below wholesale termination rates, then a strict implementation of the non-discrimination principle would call for a realignment of prices. Whether this is beneficial depends on the importance the regulator attaches to non-discriminatory prices, relative to alternative goals, such as having price correspond to a competitive outcome or realising allocative efficiency.[14]

4.2 Defining Origination and Termination Services

At the wholesale level and focussing on voice traffic, one possible division is between the services of call origination and call termination. This division would accord with the division that arises in a calling party pays environment, where the caller (originator) pays a charge for connecting with the called party, who pays nothing as a direct result of the call. The originator of the call therefore pays for the delivery of the call over the receiving party's network (call termination). Questions remain. In particular, what service are customers buying when they purchase call termination? Is call termination a bundled service that comprises coverage (as defined earlier) and conveyance of the call across the mobile network?

[13] Article 6 (a): Directive 97/33/EC of the European Parliament.

[14] See Section 7 for more discussion on how regulatory intervention might affect realisation of these goals.

The debate centres on the charging arrangements for the various services. In the calling party pays environment, should the caller pay for coverage, or should it be a separate service for which the subscriber pays? Is the caller to a mobile network purchasing the coverage that a GSM network offers when paying to have a call terminated?

Subscribers to a mobile network presumably value the coverage the network provides. Callers to a network may also value the coverage the network provides because it permits the caller to establish a link with the subscriber without needing to know where the subscriber currently is. In contrast, the subscriber originating a call knows his/her location when making the call. One might argue that this implies that callers to a mobile network may value coverage more than callers from the network. However, who values something more does not necessarily determine who pays. This approach does not help answer the question of whether, if prices were in some way cost-reflective, the calling party's purchase of call termination service would include some payment for coverage.

One viewpoint is that coverage is a service that subscribers to a mobile network purchase or acquire as a direct result of their subscription to the network (and through nothing else). Subscribers have purchased mobility and the option to make and receive calls within the network's coverage area. Under this line of argument, call termination is simply conveyance of the call because the receiving party (the subscriber) already owns the option of receiving this call (providing he/she is within the coverage area, subject to certain probabilities of call blocking and dropping).

But it does not seem certain that subscribers to the network have directly purchased this particular option. Pay-as-you-go arrangements do not involve a subscription fee. An alternative stance would be that these consumers have not purchased coverage as a stand-alone service, but rather they have acquired an option to purchase bundles of coverage and traffic if and when required. Similarly, they have acquired the option for potential callers to them to purchase bundles of coverage and traffic. Under this approach, coverage is not a service that is purchased separately. Rather it is bundled with other services to create the services that users actually buy: subscription to the network; call origination; call termination; data services etc.

The definition of a termination service seems somewhat arbitrary; neither of the conflicting viewpoints introduced above appear objectively wrong. But the latter approach has some practical advantages. Customers who go abroad and roam will not pay a separate subscription charge to the network that they are roaming on. It might therefore be reasoned that the charge for terminating a call while roaming is a charge for both coverage and conveyance. Defining domestic call termination similarly will reduce the number of distinct services that need to be costed. This study therefore considers that coverage is included in a call termination or origination bundle.

This approach contrasts to that usually taken in a fixed network setting, where access charges are not included in the costs of call traffic. The different treatment can be justified on the grounds that the costs of coverage are not driven by the number of subscribers, unlike the costs of access in fixed telephony. This seems to be a very important distinction between the two types of telephony. The costs of connecting a fixed subscriber to the local exchange (access) are directly incurred in allowing that subscriber to join the network. But the costs of providing coverage are not directly incurred in allowing a particular subscriber to join a mobile network. [15]

4.3 Roaming Services

Roaming refers to the case where a subscriber to a particular mobile network uses an alternative mobile network to obtain services, rather than the subscriber's own (home) network. The alternative network can either be in the same country as the subscriber's home network or in a different country. "International roaming" describes a situation when a subscriber uses their handset on a network in a foreign country. "National roaming" occurs when a subscriber uses a different network in their own country. This typically occurs during the roll-out phase for a new network, when obligations are often placed on incumbent operators to offer roaming agreements to new entrants so that the entrant can offer customers the same level of coverage as their more established rivals. These roaming agreements may be an important feature during the roll-out of 3G networks by new operators.

Focusing on international roaming, there are three types of calls that a roaming subscriber can make and receive:

1 *In country* calls (e.g. a France Telecom Mobile subscriber, roaming in Germany, phones a colleague in Germany);

2 *Calls home* (e.g. a France Telecom Mobile subscriber, roaming in Germany, calls home to France);

3 *Country to country* calls (e.g. a France Telecom Mobile subscriber, roaming in Germany, phones a colleague in Italy).

Calls home account for the majority of roamed calls.

In the retail market, the prices for roaming are typically part of the bundle of tariffs that the subscriber agrees to when subscribing to a calling plan. This is not a technical requirement. At the retail level, it would be possible for a mobile customer to use one company for domestic calls and another when roaming abroad. Thus there is a case for treating roaming as a separate retail service. It need not be part of a bundle of goods sold as a package. But what exactly does a customer purchase when they roam? If roaming was purchased as a stand-alone service, the customer would be paying for various network-related services but also for a variety of retail costs — billing, administration, and advertising. Therefore, it seems appropriate to include these costs in the bundle called roaming. Subscribers who only make calls in the domestic market also give rise to such costs.

[15] Besides the additional traffic generated by location updating for an additional subscriber. This may be negligible compared to BTS and site costs.

In the wholesale market competition amongst suppliers is limited to the number of network providers in each country. Moreover, the GSM Association's framework does not deal with international roaming agreements between MNOs and organisations that are not licensed MNOs. No such agreements are currently in place. Instead the MNO in the home country will act as a reseller of wholesale roaming services purchased from other MNOs. The network operators provide the network services that they also offer to their own domestic retail customers, including conveyance of calls and coverage. Therefore, the costs of roaming should include the costs of coverage. This may or may not create problems if trying to compare apparently similar services, such as call origination on the home network versus call origination when roaming. It depends on whether call origination is defined in the home market to include coverage.

Further discussion of roaming is left until Section 8 of the report, although the ideas discussed in the intervening sections are ideas that have relevance when considering the costs and prices of roaming services. For the remainder of this report, roaming services in the wholesale market are defined as call origination and call termination, with both these services assumed to be bundles featuring both conveyance and coverage. In the retail market, retail costs for things such as administration and billing are also included in the services to be costed.

4.4 Summary

We have discussed above both the importance of defining services precisely if they are to be costed individually and, in the context of the mobile industry, the difficulty in doing so. For the majority of mobile services, it is not obvious exactly what the service comprises. This is particularly the case with respect to the treatment of coverage. We have proposed that coverage is not treated as a stand-alone service that users purchase, but rather it is bundled with the services such as call origination and call termination. Amongst other arguments supporting this approach, the different treatment from the fixed network approach can be justified because coverage, unlike access in fixed networks, is not driven by subscriber numbers.

A division between wholesale and retail services is also constructive because of the strict barriers to entry at the wholesale level that do not necessarily exist at the retail level. But this distinction is not problem-free. In particular, a considered analysis of advertising by an MNO suggests that some (but by no means all) might be seen as wholesale activity. The following table summarises a possible breakdown between services that might be used.

Table 4.1: Service Breakdown Wholesale / Retail

Wholesale	Retail
Network infrastructure services provision[16]	Customer terminal subsidies[17]
Wholesale billing	Customer relations / services
Other wholesale administration	Other retail administration
Advertising and marketing	

Separating retail from wholesale services is important when we examine the costs of network services such as call origination and termination. For clarity of exposition in the following sections we focus on a simplified interpretation of wholesale that does not include allowance for advertising. Advertising benefits that accrue to wholesale could be factored in afterwards as an externality between retail and wholesale services.

A retail / wholesale distinction also provides insight on the implications of a non-discrimination obligation. The principle might be interpreted to imply that the wholesale termination charge of a SMP operator should be below its retail on-net charge. Issues surrounding the relationship between costs and prices are discussed in Section 7.

The precise definition of a roaming service is also uncertain. We propose that it does include coverage at the wholesale level, and, at the retail level, includes the services of billing administration, as well as a portion of advertising.

16 Including spectrum costs as part of the radio access network.

17 The Commission has indicated that to ensure technology neutrality between mobile and fixed networks, it believes that terminal subsidies should be treated as retail costs and not be externalised to interconnecting operators.

5 COSTING SERVICES

Summary

- There are not significant differences between the costs of terminating a call and originating a call.
- But there are significant joint costs. These mean that it is impossible to uniquely identify the cost of a service. Instead it is important to be precise about what cost concept is being used.
- Incremental and stand-alone costs of the services are very different – more so than in fixed telephone networks.

This section discusses the methods by which the cost structure of a mobile network can be defined and estimated. Identification of the cost of a telecommunications service should be separated from the issue of an appropriate price for this service. For a given set of policy objectives there may be an optimal price for a particular service that may or may not be directly related to the costs of providing that service. This idea is discussed in more depth in Section 7.

There are some important differences between fixed and mobile networks when costing services, as this section makes clear. Following on from Section 4, identification of appropriate increments for costing purposes presents new problems over those faced in a fixed network setting. Moreover, there are sizeable joint costs however the increments are defined.

5.1 Costs Characteristics of a Telecommunications Network

A telecommunications network is characterised by substantial economies of scale and scope. Economies of scale are present when the average cost of producing X units of a service falls as X rises. Economies of scope are present when the cost of producing two distinct services together is less than the costs of producing the two services separately.

When there are economies of scale, the average cost of a service exceeds its marginal cost (the cost of producing another unit of the services). If there is interest in ascertaining whether the prices charged are just sufficient to recover total costs then looking at marginal costs will not be appropriate. But the marginal cost may be of interest if a policymaker wishes to ascertain how much it will cost to provide an additional unit of the service.

Economies of scope between services are associated with joint costs. These are costs incurred in the production of multiple services that are invariant to changes in the quantity produced of one service alone. A good example of economies of scope in mobile telephony is the joint cost arising from conveying voice traffic and delivering SMS text messages over the same network. There is a substantial saving in network infrastructure when these two services are provided together.

Moreover, there is no objective way to allocate the joint costs between the two services to arrive at a cost for each service, beyond identifying stand-alone and incremental costs.

The effect of substantial economies of scope between services is to create a large difference between the incremental and stand-alone costs of the services. The stand-alone cost of providing a service would be the cost of providing the service given no other services are being provided. Consequently, this definition of the cost of a service includes all the joint costs. In contrast, the incremental cost of a service is the additional cost that arises because of providing the service given all the other services that the network provides. No joint costs are included when calculating the incremental cost of the service.

Box 5.1: Defining Cost Concepts

There are a variety of different cost concepts that economists consider. In talking about the cost of something, it is important to understand what cost is meant. It is also important to stress that these cost concepts can be analysed in isolation from prices. Section 7 discusses what relationships between costs and prices might be.

Total cost: the cost of producing all the units produced.

Marginal cost: the cost of producing an additional unit.

Average cost: the total cost divided by the number of units produced.

For a firm that produces many distinct goods or services, the following concepts might be used when looking at the cost of producing just one of the goods or services.

Incremental cost: the cost incurred providing an activity or group of activities given that all other activities are already being produced.

Stand-alone cost: the cost incurred providing an activity or group of activities in isolation

Joint cost: the cost incurred producing multiple activities that is invariant to changes in the quantity produced of one activity alone.

5.2 Joint Costs in Mobile Telephony

There are joint costs in fixed networks. For example, the access and core networks will share some trenching; these will be joint costs between the two. The proportion of joint costs between increments is generally greater if looking at smaller increments (e.g. leased lines / voice / data will share the same trenching).

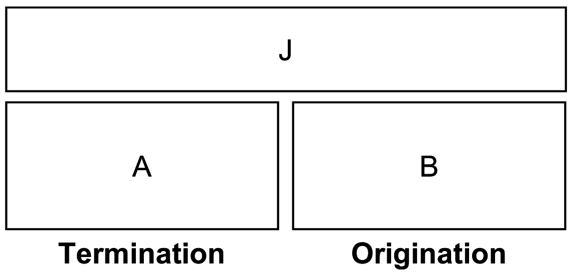
Nevertheless, joint costs are arguably more important for mobile networks than for fixed networks. Factors that give rise to joint costs in a fixed network also give rise to joint costs in a mobile network. But in addition, and more importantly, there are significant joint costs associated with providing network coverage across the country.

If services are defined as proposed in Section 4, such that call origination and call termination are bundled services that include coverage, then the costs of coverage are joint costs between these services. The same geographic coverage network is used in providing origination and termination.[18] The incremental cost of call termination does not include coverage costs; nor do the coverage costs belong in the incremental cost of call origination, or the incremental cost of SMS or WAP or GPRS. But a stand-alone network to provide just termination, or just origination, would require the coverage network. Thus coverage is a joint cost between origination and termination when these services are defined as bundles including coverage.

Similarly, if services are defined at the level of voice traffic, coverage costs still remain a joint cost to both voice and data traffic; the incremental cost of voice traffic (given a data traffic network already in place) does not include coverage costs, but a stand-alone voice traffic network would require coverage.

Note that the problem of joint costs does not go away if coverage is treated as a separate service. If coverage is viewed as a separate service to call origination and call termination, there will remain significant joint costs. The stand-alone cost of rolling out a network to provide coverage alone will feature many costs that would be included in the stand-alone costs of rolling out a network just to provide traffic.[19] For example, at the cell level many costs will be joint between traffic and coverage, e.g. the costs of the masts (although if the volume of traffic is large enough, the costs of masts over and above those necessary to provide coverage will form part of the incremental cost of traffic).

Figure 5.1: Incremental and Stand-Alone Costs



Suppose there are only two services that an MNO offers, call origination and call termination (see Figure 5.1). The number of minutes terminated and originated are, by assumption, the same. Assume that the coverage, as defined earlier, is the only cause of joints costs, which are equal to J. The incremental cost of termination and origination are A and B respectively. The stand-alone cost of termination would be A+J, the stand-alone cost of origination would be B+J. The incremental cost of the whole network – traffic and coverage – is A+B+J (also the stand-alone cost of the whole network). But just because the incremental cost of the whole network is A+B+J, it would not be correct to claim that the incremental cost of termination is (A+B+J)/2. The incremental cost of call termination is still A.

---

[18] It might be argued that a network that only originated calls would not undertake location updating and therefore would require a slightly differently designed coverage network. This may be so, but the cost difference between a such coverage network and one that allowed termination too is likely to be very small; the vast majority of coverage costs are base stations and sites rather than signalling.

[19] A mobile network that did not provide the service of coverage is difficult to conceptualise. It might be thought of as a fixed network that relies on wireless connections.

The choice of modelling technique will not change the incremental, joint or stand-alone costs of different services, although it may make it harder to identify the values for these different cost concepts.[20] For reasons already outlined, joint costs will be more significant for mobile networks than for fixed networks. Therefore, it is arguably more important that the regulator understand how significant joint costs are. The 1998 model used by Oftel to estimate costs for mobile operators found the LRIC for call termination on a GSM 900 network to be 5.3 pence per minute; our calculations suggest that the stand-alone cost for call termination implied by the model are 12.2 pence per minute – an increase of more than 125 per cent over incremental costs. Moreover, we suspect that the difference between stand-alone and incremental costs will be even more marked for many networks. For networks that have a relatively low penetration rate, the incremental costs of conveyance may be negligible. Just to provide a coverage network will provide enough capacity to carry all traffic. In this case, incremental cost per minute of call termination will be zero while the stand-alone cost per minute will be the total cost of the network divided by the number of minutes terminating.

For the purposes of pricing, it may be appropriate to “allocate” joint costs between products in a particular way. However, the costs per se of individual services are indeterminate within a range governed by the extent of joint costs (i.e. the range between incremental and stand-alone costs). In analysing estimates of the costs of different services provided by mobile operators, it is therefore essential that the assumptions concerning the treatment of joint costs are understood.

The implications of having large joint costs when comparing the costs and prices of services are discussed in Section 7.1. Below, the report discusses the different approaches used to cost services, and the difficulties with identifying a cost of a service when there are joints costs.

5.3 Approaches to Estimation of Costs

As explained in Section 4 in relation to call termination, defining the service is only the first step towards calculating “the cost of a service”. The calculation requires a judgement of what the cost refers to, particularly given economies of scope and scale. As noted above, the difference between the incremental and stand-alone cost estimates for a service depends on the size of the joint costs. If these are large, the definition of cost chosen will have significant implications for the size of the reported cost of the service.

5.3.1 Retail-minus

A commonly used methodology for costing wholesale services is to calculate an operator’s total costs and subtract from these the retail costs. This approach sometimes appeals to regulators wishing to introduce competition in the retail market. Any entrant who can provide retail services at lower cost than the network operator will be able to enter the market and compete successfully. There is no need for the entrant to build its own network – something that is not possible in mobile telephony given the limited availability of spectrum. Regulators might also be tempted to calculate retail-minus costs to set price caps on the charges existing MNOs can levy on operators

---

[20] Section 5.4 discusses different possible ways to define increments for the purposes of building a cost model.

just starting to roll-out their network, to permit the new entrant an opportunity to compete immediately.

While retail-minus methodologies might identify that, say, 30 per cent of an MNO's total costs can be deemed retail costs, there is no guidance on what the costs of the various wholesale services are. There is no reason to suppose that the cost of each service at the wholesale level is 30 per cent less than the price that the MNO charges customers (see Section 7 for more discussion on relationships between prices and costs). This is a drawback if the intention is to ascertain the costs of a roaming agreement or the costs of providing interconnection to other operators wishing to have calls terminated.

5.3.2 Fully-allocated costs

A fully-allocated cost methodology, or FAC, allocates the whole of a network operator's costs to the different services it produces.[21] Where there is a one-to-one mapping between different costs and different services (i.e. each output is exclusively produced by a unique input) it is relatively straightforward to assign different costs to different services.

Where a cost is incurred in the production of multiple services, the allocation is somewhat more troublesome. Nevertheless, it is likely that where costs are incurred in the production of multiple services, these will not solely be joint costs. Any elements of the costs that vary directly with the output of a single service can, in principle, be allocated to that service on the basis of the appropriate cost driver. In practice, the extent to which costs can be allocated to the services that cause them will be affected by the sophistication of the accounting system used. The use of an activity-based costing (ABC) system, which traces costs to services via the activities each service entails, will help in this regard. But, as discussed above, there is no single meaningful attribution of a joint cost between the different services involved when looking at their costs. Instead they will be allocated by some rule such as a proportionate mark-up over attributed costs of each service or a mark-up based on an output measure.

The need to allocate the joint costs between different services means that it is unclear exactly what the FAC of a service measures. Summing the FACs of all the services will generate the total cost of the network (in contrast to summations of the incremental or stand-alone costs of services), but the FAC for an individual service is a number between the incremental and stand-alone cost of the service that reflects the modeller's prejudices in terms of how joint costs should be allocated to different services.

5.3.3 Forward-looking long-run incremental cost

The dominant methodology for the costing of telecommunications services is the estimation of FL-LRIC. In fixed-line telephony this approach is favoured in the EU (and elsewhere, e.g. the USA and Australia) as the basis for setting cost-reflective prices for services such as interconnection.

---

[21] It is also referred to as fully-distributed cost or FDC.

FL-LRIC can be defined in two complementary ways:[22]

- It is the additional cost the firm incurs in the long run in providing a particular service (or production element) as a whole, assuming all its other production activities remain unchanged.[23]
- It is the total cost the firm would avoid in the long run if it ceased to provide the service (or production element) as a whole.[24]

An *increment* can be thought of as a finite quantity of a particular input or output. This should be contrasted with the term *marginal*, since the marginal output would refer to the last (infinitely small) unit of an output being considered. For a telecommunications network, it would seem more appropriate to analyse the costs of a specified increment of output rather than the marginal cost of a specified output (because of the substantial economies of scale in the network).

In practice the increment is often defined as all units of the given service or increment, e.g. the increment might be all call-origination minutes. The incremental cost of call origination would be the annual cost avoided by the network operator in providing every service as at present, excluding call origination. Given X minutes of call origination per year, the FL-LRIC of origination could then be divided by X, to provide an estimate of the average incremental cost of one minute's call origination. If the increment studied is access, it may be appropriate to divide the FL-LRIC of the access service by the number of customers, to obtain an estimate of the average incremental cost of providing each customer access.[25] Moreover, joint costs within the increment tend to be implicitly allocated by equi-proprtionate mark-ups.

Where there are joint costs between services within an increment and services outside the increment these would not be included in the estimation of FL-LRIC. Hence the presence of joint costs between increments implies that the sum of FL-LRICs across all increments of the network will be lower than the (forward-looking) total costs of the network.

---

22 Australian Competition & Consumer Commission (December 2000) "Pricing Methodology for the GSM Termination Service — Draft Report".

23 The long run is defined such that all factors of production can be varied.

24 This is also referred to as the decremental cost.

25 The term FL-LRAIC (forward looking long run incremental average cost) is often used to denote a FL-LRIC model in which this averaging over call minutes or number of customers is applied. However, the term FL-LRIC is frequently used in such cases.

As a consequence of this, a network operator that is only allowed to recover the FL-LRIC associated with each service it provides would incur a loss if joint costs existed between increments. For this reason, when using FL-LRIC to set prices regulators in both the EU and the US have conventionally added a mark-up to each service in order to allow an efficient operator to recover the totality of its costs. It is useful to identify which of the "costs" of a particular service are accounted for by FL-LRIC and which are due to a mark-up. Where a mark-up is included, it is not inconceivable that it will be allocated along similar lines to the method used under FAC. Note that if a mark-up is included, the calculated cost of the service reflects the modeller's prejudices in terms of how joint costs should be allocated (as under FAC).

FL-LRIC is by definition a forward-looking methodology: no regard is given to the prices paid for the various assets that are actually used in the production of services. Instead the calculation is based on the costs of a network operator using the most efficient technology to produce the same set of outputs as the operator being modelled.

5.4 Defining Increments

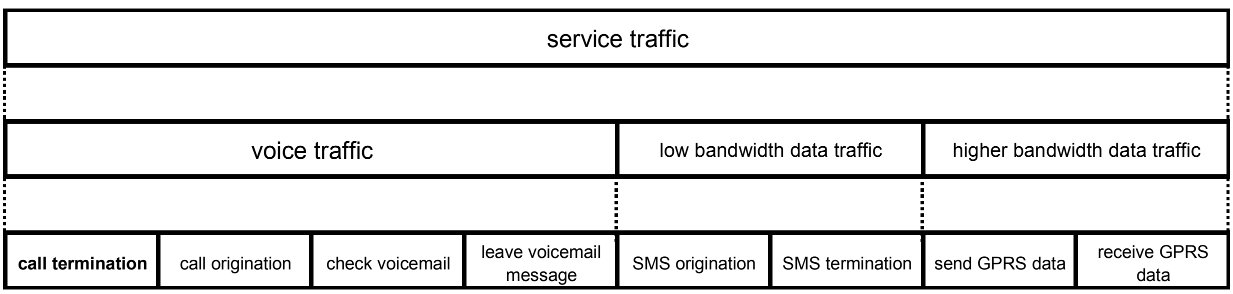
This section considers ways to define increments in the context of a mobile network operator selling services such as origination, termination and voicemail. The focus is on wholesale services. The costs of advertising and sales are not included in this analysis.

Because a principle aim of an FL-LRIC analysis in mobile telephony might be to gain an understanding of the costs of call termination, we examine the choice of increments with particular reference to this. Under a broad FL-LRIC approach, there are several ways to define the increments in a mobile network. These can be thought of as distinct levels at which to analyse services. Three potential levels at which increments could be defined at are:

- All service traffic
- Voice traffic
- Call termination

How various final services would fit into this hierarchy is illustrated in the following diagram.

Figure 5.2: Possible Increments



Having defined the increments, there remains the task of allocating costs to different services, and to units of service, within the increment. For example, if the increment is defined as voice traffic, it would be useful to estimate the cost of this increment to obtain an estimate for cost per minute of call termination. There are several ways by which costs within an increment can be allocated, e.g.:

- Use one unit of measurement within the increment to obtain the average cost of each service with respect to this measure; or
- Fully allocate costs between services within the increment based on cost drivers.

Which of these is most appropriate will depend on how the increments are defined. This is examined below, along with a discussion of the different ways in which to define increments. Generally, the starting premise is that it is reasonable to average costs of the same service across all units of that service provided by the network (so the cost for one minute of termination does not depend on *which* minute it is). But where it is necessary to attribute costs to different services within an increment, a more complex approach may be appropriate to take account of cost differences between services.

5.4.1 Subscribers

Parallel to the increment(s) of traffic-related services, we might also consider increments of coverage and subscribers. The subscribers increment would include those network-related costs that are driven solely by the number of subscribers on the network, assuming a constant output of the various traffic-related services produced by the network. On this basis, the increment of subscribers includes the cost of handsets and SIM cards (one per customer). (Since we are considering the wholesale rather than retail level, there is no need to include the costs of advertising or customer care.) Once the incremental cost is obtained, these can be allocated on a per subscriber basis by averaging the cost over all subscribers to the network.

The number of subscribers to the network will affect the required capacity of the HLR database. However, there are questions as to how much of the cost of these would be incremental to the traffic-related services provided. Both the HLR and VLR are essential for call termination. If it is the case that the entire cost of these registers is joint between customers and call termination, then these costs will not be part of either increment. Only the HLR capacity in excess of that required for the level of traffic demand for which the network is dimensioned would be incremental to subscribers.

Another candidate for incremental subscriber costs is the costs incurred in location updates. One view is that although these are essential as a pre-requisite for the provision of services across the network, the costs of location updates are not part of the provision of the final services themselves. Since subscribers automatically receive mobility, subscribers would drive the costs of location updates. The additional cost of location updates due to subscribers can be considered as the extra (annual) cost of providing location updates across the whole network to N subscribers compared to the cost of providing such location updates to one subscriber. Again, of these costs, only those that are not joint with the provision of other increments are included in the increment of subscribers.

5.4.2 Call termination increment

For traffic, one approach would be to define increments at the level of the final services that the network provides to subscribers. The increment could be one minute of call termination or one SMS message. However, to obtain a call termination cost, it seems better to consider some measure of the total cost of termination the network incurs per year and convert this to a per minute cost. The smallest reasonable increment seems to be the provision of total call termination services (perhaps to just one originating operator) in one year. Smaller increments will largely comprise joint costs.

Even looking at all call termination minutes, the joint costs between increments would be large. For each increment, FL-LRIC would essentially be composed of any service-specific network infrastructure plus the extra costs of network capacity the service uses. The latter would include any extra switches, base stations and location registers that are necessary solely because of the extra capacity required by that service. The high capacity of fibre means that in many instances the costs of fibre would not be included in the costs of an increment.

Once an estimate is obtained for the FL-LRIC of each service, call termination can be expressed as cost per minute of call termination by dividing the FL-LRIC for call termination by the total annual minutes of incoming calls to the network. Similarly for the costs of sending SMS messages, we can average the incremental cost of the SMS (sending) service across all messages sent annually from the network.

One problem with the approach described in the previous paragraph is that the link between cost causation and purported costs has been lost. The incremental cost of call termination will depend mainly on the extra traffic that the network is dimensioned to carry during the busy hour. There will be few costs caused by calls made outside the busy hour.

Box 5.2: Cost Difference Between Origination and Termination Across a Mobile Network

What is the difference in costs between call termination and call origination? Appendix A2 identifies the differences in terms of network elements used when originating and terminating calls across a mobile network. In particular, costs may vary between the two services because of the procedure for call set-up and the route the call takes across the network. Some simple calculations can provide a crude indication of the magnitude of these costs differences.

Call set-up

Both call origination and termination share a common stage of set-up communication between the subscriber's handset, the VLR and the GMSC. However, call termination requires an earlier stage of initial information exchange between GMSC, HLR and VLR (to locate the subscriber). One method for estimating the cost difference that this implies is to analyse the extreme case where we allocate the whole cost of a network's HLR component to call termination and none of this cost to origination (or other services that might use the HLR).

We consider a network serving 10 million subscribers, terminating a total of 5 billion incoming call minutes per annum. Suppose the annualised HLR cost associated with serving this number of subscribers is €5 million. Allocating this HLR cost entirely to incoming call minutes infers a cost per terminated call minute of 0.1¢. In addition to the information obtained from the HLR, call termination makes more use of the VLR than origination (treating location updating as a separate service). Allocating all the VLR costs to termination could result in an additional cost of a similar order of magnitude to the HLR costs.

Although these figures are purely indicative, it suggests that the set-up cost differential between origination and termination are small. This excludes the routing of the initial signals across the network, but these must be marginal given the low capacity required for signalling compared to voice traffic. Moreover, the estimate is biased upwards by the fact that all processing costs have been attributed to incoming calls. A better estimate might be to allocate the costs in the ratio 2.5:1 between call termination and call origination.

Call Routing

Because a mobile network operator can choose the GMSC it originates a call to, but cannot determine the GMSC it must terminate a call from, the costs of conveying outgoing and incoming calls across its network may differ.

One approach to estimate this cost difference is to assume that on average each terminated call travels through one more MSC than each originated call. If the network's GMSCs were fully intermeshed, this would be the worst-case scenario in terms of routing an incoming rather than outgoing call.

A network might be designed so that an average MSC handles 300,000 busy hour call attempts. Suppose there are 5 billion incoming call minutes per annum, we can try to convert this is to a reasonable figure for busy hour call attempts. Assuming an average call length of 2.5 minutes, this number of incoming minutes equates to 2 billion successful calls. Assuming that one in five call attempts are unsuccessful (i.e. not answered with the caller hanging up before voicemail is activated), this infers 2.5 billion call attempts per annum. Supposing that 0.1 per cent of these call attempts are in the busy hour, the 5 billion incoming minutes translates to 2.5 million busy hour call attempts.

Hence, in this example, 9 extra MSCs would be required for the 5 billion incoming minutes that the network terminates, compared to the case where these 5 billion minutes are originated by the network. If the annualised cost of each of these MSCs is €1.5 million, the extra cost to the network because these calls are incoming rather than outgoing would be €13.5 million. Allocating this evenly across all incoming call minutes suggests an additional cost for termination in the order of 0.3¢ per minute.

To this needs to be added the costs of the extra transmission stage, between GMSC and MSC. But this will be small. A cost per minute calculation for transmission is unlikely to exceed 0.1¢

Other costs

There are other costs that the network operator may well seek to recover from the prices levied on call minutes, whether originated or terminated. For example, the costs of billing, marketing, administration, even handset subsidies are all costs the MNO will need to recover. However, it is difficult to argue persuasively that these are costs that arise mainly because of call termination rather than call origination; or even that many of them are costs of a wholesale service.

In conclusion, the difference in costs between call origination and call termination in terms of the different elements in the network the two procedures use is not large; perhaps less than 1¢.

5.4.3 Voice traffic increment

At a higher level than call termination, we could define increments as voice traffic and different types of data traffic. Included in the voice traffic increment would be call termination, call origination and accessing voicemail. Data might be divided between SMS (low bandwidth) and higher bandwidth data such as WAP and GRPS.

Network elements used solely for the provision of voice traffic would fall under the voice traffic increment. In addition the extra capacity that voice traffic requires (compared to that needed if only the data traffic was carried) would be included in the increment. This would probably take the form of extra switches and more cells sites as capacity constraints in urban areas bring the need for cell splitting.

Once incremental costs are found for voice and various types of data traffic, there are two alternative methods to allocate these in order to obtain, for example, a per minute call termination cost. The first would be to find the average per minute cost of voice traffic by averaging the FL-LRIC of voice traffic across the total number of voice minutes. Note, however, that any cost differences between call origination and termination in terms of network elements used will not show up in the figures calculated.[26] This is because averaging across voice minutes does not take account of the different (per minute) requirements that different voice services may place on the network, in terms of both specific elements and capacity.

Alternatively, the FL-LRIC of voice traffic could be attributed to the various services within the increment, perhaps using an FAC approach. This approach would allow a different per minute cost to be calculated for call termination, call origination and access to voicemail. If joint costs between call termination and origination are present within the increment, some method to allocate these will be needed. Averaging across minutes simply implies an equal allocation of joint costs within increments to each minute.

5.4.4 Service traffic increment

Yet another approach would be to focus on one large increment of traffic over the network. This would include both voice and data traffic, so services such as sending SMS messages would be in the same increment as call termination.

Looking at the increment of all service traffic results in a less clear definition of what the costs are without the increment in place. In the previous cases, the incremental costs would be the difference between the costs of providing a network that did not carry voice calls versus a network that does carry voice calls; or the difference between the costs of providing a network that does not terminate calls, versus a network that does terminate calls. But if the comparison is between a network that carries traffic and one that does not carry traffic, what exactly are the costs of a network that carries no traffic?

At one extreme, the service-traffic increment would actually be the entire mobile telephone network. All services provided are related to conveying voice and data across the network; costs that are neither incremental to services nor joint between them are not network costs. An alternative is to define the incremental costs of service traffic, as the costs of the extra network elements and capacity required in order to provide the final services, additional to those required to provide coverage or part of the subscribers' increment. The advantage of this approach is that if the regulator subsequently wishes to use the model to consider appropriate prices of services, many of the joint costs will be transparent – they would be the costs of providing coverage.

Therefore, if seeking to ascertain the costs of services, the first step should be to cost a coverage network that permits a subscriber to make or receive a single call from anywhere in the network.

26 See Box 4.1.

The incremental cost of service traffic can then be obtained by removing the corresponding cost of the coverage network from the cost of a network dimensioned to provide the full volume of final services (to one subscriber). A key difference between the two networks is that the coverage network would have very low capacity because it is dimensioned for just one subscriber, only using the network for a single call.

The incremental costs of traffic would be the costs of:

- the extra transceivers/towers/antennas needed to serve the traffic on top of those already in place to provide coverage;[27]

- all the extra transport and switching equipment needed to serve the traffic on top of those necessary for coverage; and

- the costs of providing extra processing capacity for the VLRs and HLR, over that required to serve subscribers and provide coverage.

The increment may also include the costs of any additional spectrum that the MNO uses to provide 2G services over and above the costs of the spectrum used to provide coverage. For many MNOs there will be no spectrum costs in the traffic increment because the MNO had to purchase the spectrum in a single block, so the cost of the spectrum is joint between both traffic and coverage.

The cost drivers for the traffic increment will be call attempts and call minutes for voice services, and capacity for data services.

The incremental cost of traffic may be very small in countries with low population densities, since the coverage network may have sufficient capacity that little extra equipment needs to be installed. But for networks that carry a lot of traffic in urban areas this will not be true. So if networks in Luxembourg and Finland carried the same level of traffic, the incremental cost of traffic may be considerably higher in Luxembourg. This need not mean that the prices of traffic services should be higher in Luxembourg, since the costs of coverage will need to be recovered and these will be larger for the network in Finland.

---

27 The quality of service, in terms of number of dropped and/or blocked calls, plays an important role in the extra transceivers/antennas attributable to conveyance. Quality of service should be taken into account when comparing costs of networks in different countries.

Table 5.1: Assigning Major Cost Categories to Increments

Cost category	Subscription increment	Coverage increment – stand-alone costs	Traffic increment
Spectrum		Y	Possibly (if spectrum can be purchased in separate blocks)
Site – rural		Y	
Site – urban		Y	Y
TRX		Y	Y
Backhaul		Y	Y
BSC		Y	Y
BSC – MSC		Y	Y
MSC		Y	Y
Network management	Y		
HLR	Y	Y	Y
VLR	Y	Y	Y
Authentication centre	Y		

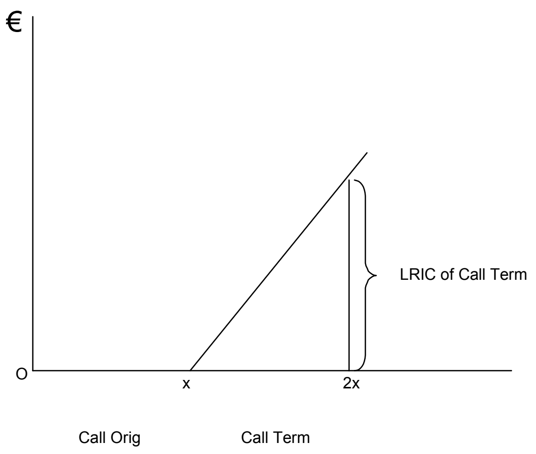
The joint costs between coverage and traffic mean that sensitivity analyses cannot say that increasing the area covered or the number of call minutes by 10 per cent will increase total costs by x per cent for all networks. Such a calculation would need to be made on a network-by-network basis, specifying the starting point.

5.4.5 Costing services given choice of increment

The costs of providing coverage, once incurred, will permit a certain amount of traffic over and above the one call that the coverage network was designed to carry at no extra cost. A rural site with a single antenna might be suitable for handling 300,000 minutes per year.[28] Provided traffic does not exceed this level in all sites, then there is no incremental cost associated with traffic and it is also correct to state that the incremental cost of call termination or call origination or any other traffic service is zero (ignoring for now the possibility that additional switches and transmission are required). But suppose total traffic was sufficiently high that the number of TRXs be doubled. The incremental cost of traffic is the cost of the additional TRXs. But how can these costs be attributed to services? The incremental cost of termination given all other services will be the entire cost of the extra TRXs (see Figure 5.3).

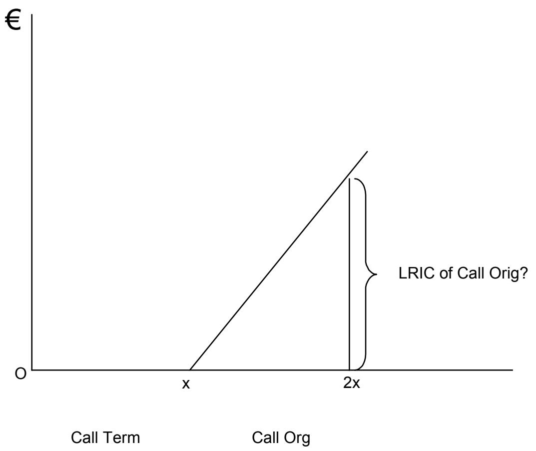
[28] The actual number will depend on assumptions about the percentage of traffic during the busy hour.

Figure 5.3: Incremental Cost of Termination



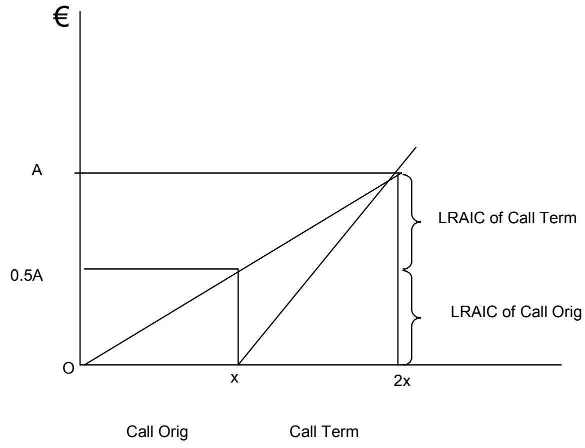
This would also correspond to the incremental cost of call origination given all other traffic (see Figure 5.4)!

Figure 5.4: Incremental Cost of Origination



But if a regulator calculated the incremental cost of call origination, and having done this then sought to ascertain the incremental cost of call termination, it would now be zero. The incremental costs of services will depend on the sequence in which they are calculated. This is potentially unsatisfactory from a regulatory perspective. The problem arises because of the problems of scaleability. It will be most acute when considering those elements of a mobile network connected with the radio interface, elements that account for over 50% of the total wholesale costs. A site with a single TRX and BTS, necessary to provide coverage, could handle many call minutes before additional equipment was necessary.

Figure 5.3: Average Incremental Cost of Origination and Termination



A solution would be to adopt the LRAIC approach, and to average the incremental costs of traffic across the various services within the traffic increment to calculate the LRAIC of the services (see Figure 5.5). LRAIC charges for the individual services would be calculated in a similar manner to the way they are calculated for fixed networks. The units of measurement for the volume of traffic for a given service will have to be converted into a unit common to all the different services using a mobile network. In many cases the obvious candidate will be Erlangs. These will then need to be adjusted to reflect the different intensity with which different services use various network elements. For example, a minute of call termination will require more switches, in expectation, than a minute of call origination. Roaming services give also rise to additional signalling, so a minute of roamed conversation implies marginally bigger capacity requirements (although of small entity) than a minute of conversation originating from the home network. This should be taken into account when costing roaming services.

Trying to ascertain the extent to which data services add to the capacity requirements of a network will be harder, a problem that is likely to become more important as 3G services are rolled out.

5.5 Recommendations if Building a Cost Model

The reason the cost model is being built will influence the appropriate way to build the model. If the regulator just wishes to know the stand-alone cost of a service, one option would be to build a bottom-up model for a network that only provides one service. In other circumstances, it might be desirable to calculate the incremental costs; perhaps the regulator is thinking about setting a price cap.

If the interest is in the incremental cost of a given service, then the best way to determine this is to calculate the costs of a network that provides that service and compare this to the costs that a network would incur if it did not provide the service. We suspect that in many cases, the regulator will want to calculate the costs of a number of services, not just one. Calculating the incremental costs of the services sequentially will produce results that are sensitive to the sequence in which the services are costed. To overcome this, we would recommend calculating average incremental costs, having first calculated the incremental costs of carrying traffic over the network.

A big difference between costing fixed networks and costing mobile networks is the proportion of joint costs in mobile networks between different services. As discussed in Section 7, joint costs have important implications when thinking about the relationship between costs and prices. Therefore, a good cost model should be transparent and clearly identify the joint costs.

One way to do this would be to estimate the costs of providing a coverage network. These costs are joint to all the services that use a mobile network and will account for the majority of joint costs. The simplest way to do this would be to use a bottom-up model that estimates how many sites would be necessary to provide coverage.

6 EFFICIENT COSTS IN A DYNAMIC SETTING

Summary

- If an MNO sells some services into a competitive market, then market forces should ensure that the operator is reasonably efficient.
- MNOs engage in not just price competition — they also compete in terms of coverage, quality and range of service. Investments are made to offer improved services in the future. This has implications when trying to cost current services.
- To calculate the costs of an efficient MNO requires judgements about how the network should be dimensioned. It is also possible that the current network topology does not minimise current costs but is sensible given future roll out plans.

In all Member States there are competing mobile operators. The form of competition between competing MNOs is not just over price. The operators will also compete in terms of the quality and range of services they offer, and the coverage. Investments are constantly being made to offer new or improved services; such investments can leave some assets economically obsolete before they are physically non-functional.

As networks have been rolled out, more subscribers have joined the networks permitting economies of scale to be realised. In the early years an MNO may concentrate on providing sufficient coverage to attract customers. Different networks with different levels of coverage and subscriber bases might have very different costs. It might be argued that the differences are largely disappearing, at least in terms of differences in coverage, as mobile operators complete their roll out. Nevertheless, differences in costs are likely to persist because of differences in the nature of the services that different operators offer. For example, some MNOs may concentrate on providing data services, which may result in a very different cost structure to that of MNOs who primarily cater for voice traffic.

6.1 Productive Efficiency in Mobile Telephony

As mentioned in Section 2.5, a regulator may have a number of objectives, potentially in conflict. The trade-off between various regulatory objectives has been important in designing schemes for the regulation of natural monopoly infrastructure in electricity, gas, rail, telecommunications and water. But it is not clear that the trade-off between productive and allocative efficiency would always apply to regulation of mobile telephony.

Suppose that a regulator decides that there is a separate market for call termination to each network and that within each, the network operator is dominant. Suppose furthermore that intervention is deemed necessary to limit the ability of operators to exploit their dominance in these markets. What type of regulation is appropriate? A key consideration would be whether the industry as whole is competitive. This would be the case if, for instance, any monopoly rents obtained in the call termination market were competed away through incentives for operators to lower prices for the other services that they provide. If firms are subject to strong competitive pressures in some areas of mobile network services (such as subscription and call origination) then we may be sceptical about whether regulation should focus at all on productive efficiency.

An understanding of the infrastructure required to provide call origination and call termination quickly reveals that there are very few network elements dedicated to providing just one of these services. For example, the same base stations, sites, switching centres and fibre will be involved in facilitating both call origination and termination. A fibre link may be simultaneously transmitting traffic that is nominally call origination and call termination. It seems very unlikely that a firm can, at the same time, exhibit productive efficiency within competitive markets and productive inefficiency in the non-competitive markets. (Concerns may remain however with relation to the incentives for innovation that would be specific to a service in an uncompetitive market, with no overlaps with competitive services.)

This analysis suggests that if concerns are not over competition in the mobile sector as a whole, but specifically in markets for particular services such as call termination, then regulation is very unlikely to be able to enhance productive efficiency. This does not imply that intervention could not be beneficial. However, it strongly suggests that any regulation should be targeted at either improving allocative efficiency or wider equity considerations. This seems to imply the need for a different type of regulation than used in fixed networks because such regulation has, to varying degrees, involved providing incentives for productive efficiency.

Nevertheless, if the mobile sector as a whole is deemed not to be under significant competitive pressures, a regulator may legitimately have concerns over incentives for productive efficiency. This emphasises the need for a thorough review of the state of competition in the mobile sector and within individual markets of the sector, before any regulation is implemented. Even if mobile termination market(s) is/are found to be uncompetitive and regulation of prices seen as the solution, whether the mobile sector as a whole is competitive will have major implications for the regulatory approach that is most appropriate for any given set of policy objectives.

In particular, the importance of modelling the costs of a hypothetical new operator building the most efficient network possible at that moment in time may be reduced. Frequently, a stated goal of regulators is to set “cost-based” prices that reflect the prices that would accrue in a competitive market. If some sectors of the mobile industry are competitive, then the current costs of an MNO are the costs that have arisen in a competitive market; these are the costs that firms in a competitive industry need to recover when setting prices.

6.1.1 Network topology

The optimal forward-looking network topology will depend on the expected technological progress. For example, there is a distinction between the optimal network designed solely to provide 2G services for the next ten years and the optimal network design to provide 2G services for three years with 3G services taking a greater and greater precedence in the next seven years. Given that only 2G services are currently provided, re-optimising the network to allow for the introduction of 3G in three years would increase the current cost of 2G services since the network design is likely to differ. There are economies of scope between a 2G and a 3G network. If there were no economies of scope between the two, the design of an efficient 2G network would be independent of whether the operator was planning on introducing 3G.

Allowing 3G to influence current network design (which raises costs to current 2G users) cost savings will accrue in the future, potentially of benefit to both 2G and 3G users. The apparent cross-subsidy is not necessarily between 2G and 3G, but between current and future users. Economies of scope cannot be objectively attributed between the beneficiaries, so we cannot determine the extent to which the cost saving falls on 3G users compared to future 2G users.

To derive the notional efficient current 2G cost structure we should not take account of the move towards 3G because of this inter-temporal cross-subsidy. Operators who decide to develop 3G services have implicitly concluded that the net revenues from such a strategy are positive, i.e. the costs are less than the extra revenues that will accrue as a consequence of rolling out the service. Nevertheless, if the network being modelled is scorched-node, some allowance for network design that will (in the future) introduce economies of scope between 2G and 3G might be a practical approach. This would more reasonably approximate the network built by an efficient 2G operator with 3G interests.

6.2 Dimensioning the Network

The cost of a network will depend on the capacity the network is designed to carry. A decision may be taken that the current network currently has surplus capacity. If the purpose of the exercise is to calculate the costs of an efficient operator, then the costs of providing that excess capacity should not be included.

6.2.1 Quality of Service

When trying to calculate incremental costs, information is required on the total cost of the network, and the counterfactual of what the costs would be if the network had been designed to not carry the incremental traffic.

Suppose the increment is call termination minutes. Then if the network did not carry those call termination minutes, what quality of service would the network operator have designed the network to meet? How many calls would have been blocked or dropped? Assume that the actual network, which includes X million call minutes terminating, blocks Y per cent of calls and drops Z per cent (both terminating and originating). Then one argument might be that if the network did not have to terminate calls, it would be designed so that the same percentage of originating calls are blocked (Y per cent) or dropped (Z per cent).

An alternative argument might be that if call termination was not provided by the network, the chosen network design would block Y* and drop Z* calls. If Y* and Z* are not equal to Y and Z respectively, the network's design has incorporated costs (or cost savings) that reflect the quality of service preferences of callers to the network.

The problem is that the same quality of service is provided regardless of whether a call originates or terminates on the network. The MNO will have dimensioned the network to offer a quality of service that is a weighted average of the preferences of all users of the network, including those terminating calls on the network.[29] Consequently, if there were no calls to terminate, the network's capacity would be such that the quality of service reflected the preferences of subscribers to the network. The incremental cost would then be the difference between the cost of a network that includes X call termination minutes, Y per cent blocked calls and Z per cent dropped calls and the cost of a network that includes no call termination minutes, Y* per cent blocked calls and Z* per cent dropped calls. Notice that this approach implies that two networks with the same geographical coverage and number of call minutes to terminate (and originate) will not necessarily have the same incremental costs, since it is possible that the qualities of service will differ.

But is there an efficient quality of service? Some might argue that there is an efficient quality of service that should be offered to people seeking to have calls terminated on a mobile network. If the MNO decides to offer a higher (or lower) quality of service that is a marketing decision. Consequently, the costs (or savings) of that decision should be borne by those for whom the marketing decision was targeted. The fact that callers to the network have benefited (or not) from the higher (lower) quality of service does not mean that these are costs that they have imposed on the network. Under this approach, the incremental cost of terminating traffic might be calculated as the difference in costs between a network that terminates X call minutes and a network that terminates no call minutes, where costs are estimated assuming that both networks provide the same "efficient" quality of service. The costs for these two networks might have to be based on hypothetical networks.

Three possible ways to treat quality of service when measuring the incremental cost of call termination have been described. There are other approaches. The suitable approach depends on the reason for calculating the incremental cost. If it was purely a costing exercise, then the second approach seems appropriate — what is the incremental cost of a call termination service, taking into account the fact that the quality of service that the network provides is likely to differ depending on whether or not the service is in place. However, the practical difficulties of estimating what the quality of service would be in the absence of call termination means that the first approach — assuming a constant quality of service — also has merit.

If incremental costs are being calculated with a view to regulating charges, then there is a policy debate concerning who should pay for quality of service under calling party pays. This debate is independent of the costing approach undertaken. It would need to be addressed if the increment is defined as all voice traffic or all traffic; or if seeking to calculate FAC since a decision will be necessary concerning how to allocate costs that improve the quality of service that the network offers.

Competition is not confined to price competition in mobile telephony. It is also possible for MNOs to compete in terms of the quality of service (defined as calls blocked or dropped), the range of services, and the area for which they offer coverage. Such competition is desirable. But it can complicate attempts to estimate the costs of services, however defined. If prices are regulated for

---

29 The weighting for some services may be low, even zero. This is an empirical question.

some services, such as call termination, the decision concerning which costs can be recovered from prices for call termination, including costs incurred improving quality of service, will have an impact on the behaviour of MNOs.

6.2.2 Forecasting demand

A more standard problem with dimensioning a network is working out future traffic loads that the network will need to be dimensioned to carry. In a regulatory setting, where costs are wanted to set prices, a circularity occurs. The price set will affect quantity demanded, and therefore the optimal size of the network. Yet the size of the network will affect the estimated costs, and therefore the price set. In theory, it is possible to calculate a network for which the estimated cost coincides with the price that gives rise to demand for a network of that size. In practice, this is rarely done.

Instead, the typical approach is to forecast how demand will grow based on trends, with little consideration given to the relationship between price and quantity demanded. The appropriateness of this will depend on the elasticity of demand. If demand is not especially responsive to price this may be reasonable.

Arguably, forecasting demand for a mobile network is harder than the equivalent exercise for incumbent fixed networks, for which most costing exercises have applied. The traffic that a mobile network carries is perhaps more variable than it is for incumbent fixed network operators. If an MNO is to respond to an increase in demand for its services, it needs to have some surplus capacity. This has implications when determining the optimal dimension of the network. If the costing exercise is for the purposes of setting cost-based prices, then the regulator might want to consider the efficient level of surplus capacity each network should have if competition between networks is not to be muted by capacity constraints. If all the MNOs had no surplus capacity, then there would be reduced competition between them to attract the traffic of MVNOs, ISPs, or new entrants seeking roaming agreement during their roll-out period.

Another consideration is whether different prices should apply to different networks. Should lower call termination charges apply to a larger network than a smaller network? That is a policy decision. It does not affect the costs of a mobile network, except to the extent that the size of the network may be influenced by demand that in turn depends on price. But the costs estimated for a service such as call termination are likely to depend on the size of the network modelled.

6.3 Attributing Costs in a Dynamic Environment

The dynamic environment can also have implications if trying to assess costs at a point in time — minimising costs today may not be the same as minimising costs over a period of years. The technologies associated with mobile telephony are evolving rapidly. This has implications if seeking to estimate current costs of the network. When amortising assets, the distinction between economic and physical asset lives is a potentially important one; assets can become obsolete long before they cease to be functional.

6.3.1 Current costs

A commonly approach for estimating current cost is the Value to Owner Convention. This defines current cost as the lower of replacement cost (RC) or deprival value (the greater of either selling the asset or using the asset). This can be written as

$$\text{Min [RC, max [NRV, NPV]]}$$

The replacement cost measures the cost of replacing the existing asset with another asset of similar performance characteristics; NRV is the net realisable value, the amount that would be obtained by selling an asset; and NPV is the net present value, the sum of discounted cash flows that an asset is expected to generate during its lifetime.

In most cases, the issues surrounding current cost adjustments are the same in either a mobile or a fixed setting. (To the extent that mobile network infrastructure is typically newer than fixed network infrastructure, the divergence between historic and current costs may not be as large in mobile networks.) Current costs in principle reflect the true economic cost of an asset at a point in time, but suffer drawbacks in practice because the calculation is not straightforward. In particular, the NPV of an asset may be difficult to calculate, as discussed in the context of economic deprecation below, and may give rise to a circularity between regulatory actions and estimated costs.

One interesting case of particular importance in mobile telephony is the appropriate valuation of spectrum. Since the spectrum for 2G licenses cannot be traded, the NRV is zero. It could also be argued that the replacement cost is infinite (since it is impossible to purchase a similar asset), except on the rare occasion when further spectrum licences are offered for sale (e.g. 3G auctions). Even in these instances, spectrum is typically sold in large blocks that may be inappropriate for the replacement of 2G spectrum alone (MEA adjustments would be necessary). Furthermore, such sales are infrequent and therefore not helpful for calculating current costs on an ongoing basis.

The substantial barriers to replacing 2G spectrum mean that it is not appropriate to use replacement cost for defining the current cost of spectrum. Hence, using the formula above, the current cost of spectrum is the greater of either zero (which is the NRV) or the NPV of the spectrum.

However, there is a very important circularity that arises if the cost of spectrum is based on NPV. Besides being the asset that facilitates the radio interface in a mobile network (MS to BTS), spectrum is a barrier to entry in the industry (at the wholesale level). If a government has monopoly rights over the sale of spectrum and prevents spectrum trading, the sale or provision of a limited number of blocks of spectrum is effectively the sale of entry into the industry. Because spectrum is vital for each MNO to operate, who has spectrum determines who is in the market.

Given this factor, how much is spectrum worth to an MNO at a point in time? The answer is that it is the expected future profits available to the MNO in question (after allowing for return on capital), over the life of the spectrum rights. The formula above therefore implies that if the future expected profit stream is greater than zero, the current cost of spectrum is the expected future profit stream (allowing for return on capital). Thus, any MNO will be found to be earning normal profit at a point in time if the cost of spectrum is based on correct current costs. This circularity arises because rights to spectrum, under current arrangements, are exclusive rights to operate in the industry; these rights are correctly valued at the future rents that can be extracted from operation in the industry.

If the purpose of estimating costs is to assess the levels of competition in the mobile sector spectrum should not be costed using current costs. An alternative method is to base costs on what the MNO paid or pays for the spectrum (if anything). This could either be a yearly fee or a one-off price that the GSM operator is / was charged for its spectrum. Although these are not the current costs of the spectrum (as defined above) such an approach could be justified as these fees are the actual costs that the network incurs directly through its spectrum rights.

6.3.2 Asset lives

Many of the costs for telephone networks are capital investments that need to be amortised over a number of years. In a competitive environment, the correct way to amortise costs is to use e*conomic depreciation.* This represents the reduction in the value of an asset from one year to the next. However, it is difficult to assess ex ante because in the majority of cases we cannot accurately predict the value path of an asset over time. For example, it will be necessary to take a view on what technological advances might take place and how these advances might affect an asset's value. Furthermore, in a market that is evolving rapidly and unpredictably, it will be difficult to determine NPV because demand for services that use the asset may be very uncertain. Nevertheless, it is still possible in principle to use this approach for estimating the annual depreciation charge.

For all annualisation methodologies the assumed life of an asset will be important. Various factors will influence the asset life that should be assumed. For many assets, the economic lifetime of the asset will be less than its physical lifetime, because of rapid technological advances and because of rapid growth in demand.

Technological progress leads to the introduction of new assets that have efficiency advantages over earlier designs. The availability of superior or more efficient technology tends to reduce the value of existing assets (a good example is that of mobile handsets).[30]

Another reason why the economic lives of assets may be short in a mobile network relates to problems with scaleability. There is a limit to the extent to which many assets in a telecommunications network can be upgraded to handle extra capacity. At some point the asset will need to be replaced outright. The rapid growth for mobile services has made this problem particularly relevant in mobile telephony. Optimal network planning involves a judgement on likely future demand. With the benefit of hindsight it is possible to argue that some of the assets purchased in earlier years should have had a greater capacity. But at the time the purchase

[30] Note that it is possible for new technology to increase the value of existing assets. It could be argued that the development of the Internet has increased the potential value of telecommunications networks.

decision may have been optimal. Would an MNO have been ex ante efficient to build a network anticipating current levels of demand?

6.3.3 Cost of capital

The weighted average cost of capital (WACC) of a firm is equal to the average of the cost of debt and equity finance to the firm; weighted according to the proportion of debt to equity in the firm's financial structure (the gearing ratio). Gearing ratios are company specific, and will depend on previous financing decisions. A firm could reorder its finances if it was cheaper to raise capital by having a different gearing ratio, although this will not be costless. The current cost of capital for the operator will reflect its current gearing ratio, and be the minimum of the WACC given the company's current gearing or the WACC of an alternative gearing plus any costs incurred rearranging the proportion of debt and equity. This may not reflect the costs that an "efficient" operator would incur building a hypothetical network since the efficient operator would have the efficient mix of debt and equity rather than having to switch to this mix.

When calculating the costs of services, the standard approach is to estimate a cost of capital and apply it to both the firm that produces the service and the counterfactual of the firm that does not produce the service.[31] This is not correct. For example the gearing or the equity beta may be very different under the two scenarios. But the practical difficulties with seeking to calculate two WACCs are sufficiently large that it would be difficult to have confidence that the differences in the two estimated costs of capital were statistically significant.

The cost of debt is the sum of the economy's risk free rate of return and a risk premium that investor's apply to the network operator's business to reflect the probability of default. The risk-free rate is often proxied by the yield on government debt. The choice of maturity depends on the horizon over which the risk-free rate is to be calculated. The appropriate maturity might be one that corresponds to the average life of the operator's assets. The correct interest rate to use if looking at the current costs of the network would be the current rate on the chosen bond. Using averages of past bond yields does not produce estimates of the current costs of the network, although in a regulatory setting where prices will be fixed for a period of time the approach may have merit.[32]

The debt-risk premia of mobile companies has increased recently. The risk of default is perceived to have increased. For example, the S&P credit ratings for BT, Deutsche Telecom and France Telecom have all fallen in the last year.

The equity-risk premium is the return required on equities over and above the returns on risk-free bonds. Its calculation is contentious.[33] A CAPM methodology relates the equity risk premium to the correlation between returns on equity in the operator's business and the return on the market as a whole. The higher this correlation is, the larger is the cost of equity since investors demand higher returns for holding these shares.

---

31 An incremental costing approach, described in more detail in Section 5.

32 For example if it were assumed that interest rates are mean-reverting.

33 Arguments about the size of the equity-risk premium are not specific to mobile telephony. This report does not discuss them further.

Although the cost of capital is forward looking, in practice estimates of the cost of capital typically depend on some historical data. For example, beta is typically estimated using historical data. The beta measures the volatility of returns on operators' shares compared to the returns on equities generally. The changing fortunes of telecom equity prices in recent years also means that the estimates of beta may be very sensitive to the historical data used to estimate beta. One might argue that the movements in equity prices in recent years reflect mobile operators decisions to invest in 3G telephony. Therefore equity price movements of mobile operators in recent years are not the right benchmark to use when calculating the cost of capital for a GSM network operator.

For a diversified company, the question is whether the cost of capital invested in a mobile network should reflect the cost of financing that applies to the company as a whole, or the cost of financing that would apply if the firm only provided 2G telephony. It seems more appropriate to base a mobile network's cost of capital on the cost that would be faced by a firm that had no other operations. Costs should reflect the real economic risks involved in providing GSM services in a single country — other activities that the parent company undertakes should not affect the cost of capital. However, data considerations may prevent these costs being accurately calculated. The bias of not calculating the cost of capital for a 2G operator in a single country is unclear. It is likely that expenditure on 3G licenses has increased the costs of capital for mobile operators due to higher perceived default risks. But the impact on the cost of capital of an operator having networks in a number of countries is less clear.

The role of regulators will inevitably have an effect on the cost of capital. This is true whether or not the regulators intervene. The market's expectations about future returns will include assumptions about the likelihood of regulatory intervention and the effects such intervention will have on returns. The recent actions of the EC DG Competition and the ongoing regulatory reviews in a number of countries (see the Annex to this report) will be factors investors consider when deciding whether to invest in mobile telephony.

The actions, or inaction of regulators, and differing degrees of competition in the various member countries mean that there may be different costs of capital facing otherwise similarly sized mobile operators. Estimated annual costs could differ significantly, even if it is assumed all operators have access to global capital markets.[34]

6.4 Regulatory Implications

The main point to stress from this section is that if sectors of the mobile market are competitive, then competitive pressures should force companies to realise productive efficiency. This is arguably different to the fixed network settings regulators have traditionally regulated, where there is an incumbent with legacy equipment and limited competitive pressures.

Consequently, undertaking an exercise to calculate the costs of a hypothetical new entrant into the industry may not be as important as it has been in the fixed network environment. In most

[34] Many studies looking at the cost of capital estimate the equity risk premium by looking at the correlation between the equity price of the firm and returns on a leading national share index, rather than looking at returns from a global capital market.

cases, looking at the current costs of assets in use in an operator's network and calculating annualisation charges appropriately will suffice.

We recommend that cost models should cost a network providing the quality of service that the actual network provides. This will be the cost of that network (this assumes that the regulator is concerned about how prices relate to the network's own costs). Moreover, in a competitive market it is unlikely that all MNOs would offer the same quality of service. There is not a single quality of service that all efficient operators would provide.

Looking forward, attempts to cost individual services provided over a common network will be complicated by the roll-out of 3G networks. The current costs of an MNO's network could still be calculated. But attempts to isolate costs of individual services will be harder. 3G spectrum costs are not costs incurred to provide voice services (unless an operator can show that it was more efficient to purchase the additional spectrum than to split cells in the 2G network). In many other cases, it will be harder to isolate the costs that relate to providing 2G services. The network topology will be designed to minimise the costs of providing all the services that the network carries. This is likely to differ to the costs of a network just providing voice traffic (or to one providing data traffic). The cost of capital will also differ.

7 COSTS AND PRICES

Summary

- In the presence of joint costs, it is not sufficient to look solely at costs to explain price differences.
- In competitive markets prices should be less than stand-alone costs, but by how much will depend on the demand characteristics of the different services that share the joint costs.
- It would be misleading to concentrate on a snapshot comparing prices and costs at a single point in time.

This section discusses the relationship between prices and costs. The previous sections identify problems with attributing costs to services. There are differences between the network elements used to terminate and originate calls, but these differences are small, as Box 4.1 illustrates. But, for reasons outlined in Section 7.1 below, this in itself is not a reason to expect prices to be similar in a competitive market (or non-competitive market). To focus on costs alone to explain price differentials is mistaken.

Some other difficulties with comparing prices to costs are also identified. Section 7.2 looks at the problems of identifying a single price for a service, and discusses the rationale for two-part tariffs and competing tariff plans from a single operator. Section 7.3 discusses how prices might compare to costs when looking at the dynamic environment, and Section 7.4 tackles international price comparisons.

7.1 Pricing with Large Joint Costs

This report has already identified that there are a number of relevant cost concepts. If there are large joint costs, as identified in this study, then the difference between the incremental and stand-alone cost of a service can be large. In this case, it is quite possible for the prices of two identical services to differ substantially and yet still be "cost-reflective" (the cost of a service lies between incremental cost and stand-alone cost, but nowhere more specific). The absence of a technical explanation for large cost differences between call termination and call origination does not mean that the prices of these two services must be the same in order for these prices to be cost-reflective.

7.1.1 Pricing in competitive markets with joint costs

The furthest we can consider by looking at costs alone is the extremities of the incremental and stand-alone costs of a service. The incremental cost of a service can provide information on whether an incumbent is cross-subsidising a service subject to competition. This would be the case if revenues earned providing a given service were below the incremental cost of the service. An examination of incremental cost is therefore useful if a firm faces regulatory obligations not to cross-subsidise. However, cross subsidisation of this kind is not necessarily inconsistent with pricing in a competitive industry. For example, handset subsidies are evidence of pricing a product below incremental cost, but are the subsidies the result of lack of competition?

The stand-alone cost of a service might be of interest if the focus is on establishing whether the industry is competitive.[35] In the absence of barriers to entry, a firm should not be able to earn revenues in excess of the stand-alone cost (including a suitable return on capital employed) over a prolonged period of time. More correctly, where there is significant risk in an industry, the ex ante expected returns should not exceed the stand-alone cost. If expected returns exceed stand-alone costs, entrants should be attracted to enter the market and compete with the incumbent(s) in providing this service. If the price of any service is persistently above stand-alone cost, it is a clear sign that competitive pressures on that service are absent.

In competitive markets firms are unlikely to recover all joint costs on one service alone (e.g. price that service at stand-alone cost) or to recover joint costs by spreading them equally across all services produced.

Instead, in a competitive market, the prices will be set depending on the demand for the different products. Producers seek to price discriminate, selling each product that consumers are willing to buy, at the highest price that the market will bear. The total revenues that they raise will not exceed their total costs in a competitive market. If they did, new entrants would enter the market. For example, Levine (2001) observes that butchers sell different cuts of beef at different prices per kilogram, depending on demand (as well as supply).[36]

Relating to the mobile industry, if the industry is competitive there should be no expectation that all calls are priced at the same level. Instead, the check to see that the industry is competitive would be to compare total revenues and total costs. The evidence that provision of an individual service is uncompetitive would be to show that, for example, call termination could be provided as a stand-alone service at lower cost than the revenues operators are earning from this service. If the regulator's goal is to ensure an outcome that accords with a competitive market, the best action will be to do nothing if prices are below stand-alone cost for all services and total revenues do not exceed total costs (after allowing for a suitable rate of return on capital).

7.1.2 Ramsey pricing

A regulator's goal may be to realise allocative efficiency, rather than attempting to replicate a competitive market's outcome. If this is the case, then it is still possible that the price of different services should differ, even though their incremental costs are the same. Assuming that the regulator believes that cost recovery should be possible, an operator will need to price at least some products above the level of incremental costs. The efficient recovery of joint costs requires that products with less elastic demand should bear higher prices (above marginal cost), an idea

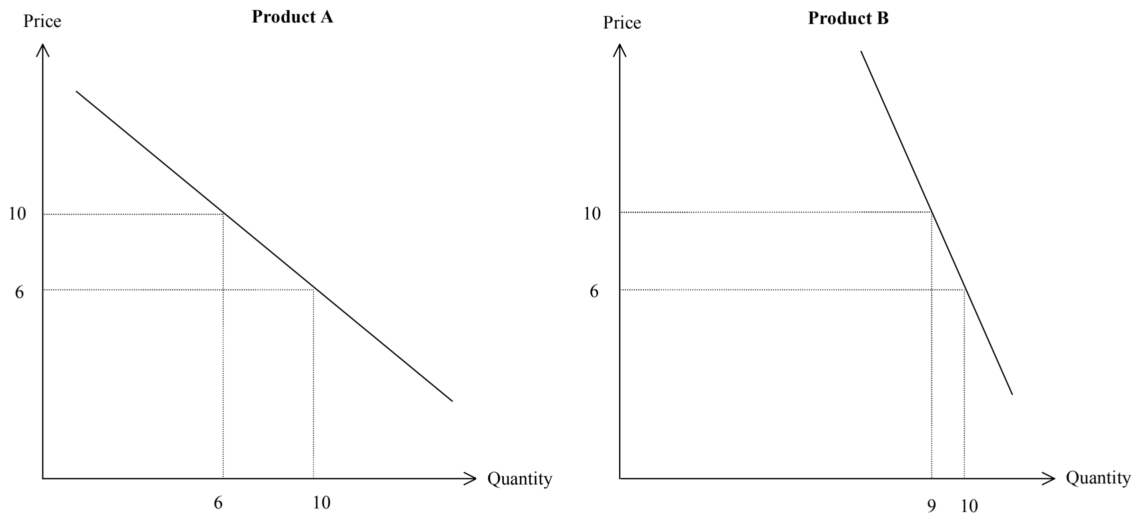
[35] This can be thought of as a necessary but not sufficient condition for pricing within a competitive industry.

[36] Levine, Michael (forthcoming) "Price discrimination without market power" *Yale Journal on Regulation.*

often identified with the work of Frank Ramsey in the 1920s.[37] The corresponding efficient prices are therefore termed Ramsey prices.

For the vast majority of products, an increase in price will cause a reduction in demand. But this effect is not uniform across all products. A 10 per cent increase in the price of one product may have a smaller impact on volume sold than a 10 per cent increase in the price of another product. The response of demand for a product to a change in its price is termed the (own-price) elasticity of demand. For a given price increase (decrease), services with a higher elasticity of demand will show a larger decrease (increase) in volume sold. This is illustrated in the diagram below.

Figure 7.1: Response to Price Changes Given Different Elasticities of Demand



Consider the case where the price of both products is 6. At these levels, demand for products A and B is 10 units each. When the price of product A is increased to 10, demand drops by 4 units. By contrast, when the price of product B is increased to 10, demand only drops by 1 unit. Product B has a smaller elasticity of demand than product A:[38] the price can be raised with a smaller effect on demand than for product A.

If a regulator were to set Ramsey prices for a firm facing joint costs, it would involve marking up prices above marginal cost (or, under uniform pricing, average incremental costs) in inverse proportion to the elasticities of demand for the various products associated with the joint costs. The discussion above suggests that this could be justified on efficiency grounds.

The problem of recovering joint costs applies when considering services in fixed as well as mobile networks. But arguably joint costs are more significant in mobile telephony. Therefore, if demand sensitivities differ between services, the divergence between Ramsey prices and prices that reflect an equi-proportionate mark-up of joint costs is likely to be greater in mobile telephony. If

37 Ramsey, Frank (1927) "A contribution to the theory of taxation", *Economic Journal,* 37, 47-61.

38 Starting from the case where both are priced at 6.

the demand elasticities differ between call termination and call origination, it would not be efficient for the two services to have the same price.

This does not mean that the current prices for call termination and call origination, or for roaming services, are necessarily at the efficient level. Instead, the discussion stresses the importance of looking at demand-side conditions as well as costs when seeking to determine the efficient prices for a multi-product firm, such as an MNO, producing goods for which joint costs are significant. Allocative inefficiency arises where consumption is reduced below the levels associated with marginal cost pricing. Suppose marginal cost is equal to 6. Then any price above 6 leads to allocative inefficiency. But Figure 7.1 shows that, for a given increase in price, this inefficiency is lower for product B (because the drop in quantity demanded is less). Since firms will only remain in an industry if cost recovery is possible, prices will need to be set in order to recover any joint or fixed costs. Not all prices can be set at marginal or incremental cost. Recovering more joint costs where demand is inelastic minimises the distortion in allocative efficiency that results from the need to price some goods or services above marginal cost in order to recover joint costs.

Whether prices for mobile services such as call termination are at the efficient level requires an understanding of the extent and scope of competition in the mobile telephony industry as a whole, as well as within individual markets for services. If competition is absent at either level, there is potential for inefficient pricing, even if this is Ramsey pricing from the point of view of each firm, if the firm faces strong competition in some markets and weak competition in others. This is because the presence of competing firms affects the elasticity of demand any one firm faces. Differences in competitive pressures in two markets will affect the relative elasticities of demand the firm faces in those two markets, causing it to price away from the efficient Ramsey levels. This is discussed in the following section.

7.1.2.1 Ramsey prices for a firm

Multiple firms within a market will create a difference between industry demand and the demand an individual firm faces. In markets where a firm faces competition, the demand for the firm's services will be affected by the supply-side strategies of the competing firms. One way to consider this is in terms of the "residual" demand a firm will face, based on assumptions of the strategies other firms will follow. In particular the presence of competition will increase the elasticity of demand any firm faces: as its price increases, consumers — besides being able to stop buying the service altogether (the monopoly case) — have the additional option of buying the service from an alternative supplier.

As competition increases in a subset of markets that the firm sells in, the Ramsey price structure (for the firm) will entail a greater proportion of costs being recovered in those markets not subject to competition. This is simply because the demand the firm faces becomes more elastic in the markets that become more competitive.

An important observation can be drawn from this. If the industry as a whole is competitive, regulation of one firm cannot improve allocative efficiency while still allowing the firm to breakeven; each firm is already pricing at efficient levels (taking the actions of competitors as given). Nevertheless, firms may not be pricing at the most efficient levels from the industry perspective.

In mobile telephony, it is likely that the demand for call origination, at the firm level, is more price sensitive than the demand for call termination. Consumers can choose which network they originate a call on; the network the call is terminated on will depend on the network that the receiving party uses. Therefore it is efficient for the individual MNOs to price call termination above call origination rates; call termination is the service from which they can recover joint costs most efficiently. If the mobile industry is competitive, then there is no advantage in regulating the prices of a single mobile operator.

7.1.2.2 Ramsey prices at the industry level

If the regulator intervenes in the mobile sector with a view to setting price controls for all firms then allocative inefficiency would be minimised by setting Ramsey prices at the industry level, i.e. based on demand characteristics at the industry level rather than those that individual firms face. For example, if the (industry) elasticity of demand in all termination markets is the same as the elasticity of demand in all origination markets, this would imply that the mark-up applied to termination and origination should be the same.[39] The Ramey price structure at the industry level is not affected by the degree of competition within each market because only industry demand is considered.

Therefore, even if the sector as a whole is competitive and firms are earning zero profits, the Ramsey prices that a regulator would try to set, in seeking to attain allocative efficiency, would be different to those that individual firms would choose if some markets are more competitive than others. Moreover, the greater is the disparity between the levels of competition in the different markets, the further away are the prices firms charge from those implied by the industry level Ramsey price structure. Intervention could in theory increase allocative efficiency in cases where this disparity is large, because the price structure that minimises allocative inefficiency in the industry as a whole would not be chosen by competing firms.

7.1.2.3 Further issues

Efficient Ramsey pricing does not just rest on the own-price elasticities of demand of the services being considered. Where demands for different services are interrelated, this will feed into Ramsey prices. For instance, if a decrease in the price of one service leads to an increase in demand for another service this will affect the efficient price structure. Specifically, the price of the former service will be marked down to take advantage of the positive influence this has on the sales volume of another service on which joint costs are recovered.

An example of this effect is handset subsidies. If subscribers to a mobile network paid the incremental cost of their subscription this would include the full cost of the handset. Does this mean that handset subsidies are inefficient or incompatible with a competitive industry? Not necessarily. If by lowering the costs of subscription, the network increases the demand for other

---

39 Ignoring the effects of any cross-elasticities of demand.

services on which joint costs can be more easily recovered, such as voice traffic, this could be more efficient than pricing handsets at cost.

It is the difference in demand between different products that renders Ramsey pricing more efficient than, say, an equi-proportionate mark-up. But demand varies not only between services, but also with other factors such as:

- the time of day or day of week;
- the particular user / groups of users of the service;
- over time (e.g. between mobile users today and mobile users next year).

If prices can be varied to take account of these demand differences, then, under the general Ramsey principle, efficiency can be increased. If the goal is to maximise allocative efficiency, allowing firms to price discriminate will permit a more efficient outcome.

7.1.2.4 Practicality and appropriateness of Ramsey pricing

Two objections are generally raised against a system of Ramsey pricing.

First, since demand elasticities are notoriously difficult to estimate with any degree of accuracy, Ramsey pricing has not been seen as an approach that can be applied in practice. This seems particularly true in mobile telephony where an empirical study to calculate industry demand appears very problematic, especially given the complex tariff structures and rapid innovation seen in the industry.

Nevertheless, if a system of cost-orientated prices is to be implemented, allocation of joint costs to different services can be based on Ramsey principles even if an explicit calculation is ruled out. If we think that industry level demand elasticities do differ significantly between different mobile services then not knowing the exact elasticities does not mean that equi-proportionate mark-ups are the only option. Some understanding of which services are likely to have the least elastic demand (at industry level) would suggest that these should recover more of joint costs than other services. Also, where the price of one product is thought to significantly affect demand for another, this may significantly influence the efficient price structure. The exact Ramsey prices may not be known (and indeed will change over time) but efficiency may be increased by a move someway in the right direction. Conversely, if elasticity of demand is generally thought to be similar across different services, then equi-proportionate mark-ups can be justified.

Secondly, equity considerations may render Ramsey prices unacceptable. Ramsey prices for a set of services are defined as the prices that minimise allocative inefficiency, given the constraint that the firm(s) producing these services must recover the totality of costs. Whether these prices are welfare maximising (subject to the same constraint) will depend on equity preferences, in particular over the distribution of income between different groups of consumers, and over time.

It is wrong to claim that Ramsey prices inherently maximise social welfare (under the constraint that firms recover all costs). This shows confusion between the concepts of allocative efficiency and welfare. The former is an objective measure, whereas different parties can legitimately hold different views on what constitutes welfare. The point is that we should be clear about the relationship between the two concepts. In the case of mobile telephony the difference between the two may be much less important than in other industries. The extent of divergence between allocative efficiency and welfare maximisation in mobile telephony can be approximated by considering whether £1 is worth more in terms of welfare in the pockets of users of different mobile services.

For example, is £1 in the pocket of a subscriber to a mobile network worth more, less or the same in terms of welfare than £1 in the pocket of someone calling a mobile network? In cases where there is thought to be little welfare difference between the two, Ramsey prices are consistent with welfare objectives. However, if it is felt that £1 in the pocket of a typical caller to a mobile network is worth significantly more (or less) in welfare than £1 in the pocket of a typical network subscriber, then Ramsey prices are not optimal.

7.2 Relationship between Tariff Structures and Costs

A conclusion from the previous section is that it is not necessarily the case that price differences between two services should correspond to differences in costs between two services. The next two sections look at the prices of one service, call origination, to illustrate some further problems with seeking to compare costs and prices. The market for call origination is generally considered to be competitive, with MNOs competing in terms of price (both levels and tariff structure design) and quality of service (in terms of coverage, capacity and range of services).

Table 7.1 presents the variety of tariff structures available to would-be Vodafone UK billing customers in June 2001. The different payment plans have different prices for different services, include varying amounts of free minutes, and fixed payments that are independent of the number of call minutes that the user ends up using. Looking at these various packages, what is the retail price of national peak calls? Is it 10, 15 or 34 pence per minute, or something different again?

Table 7.1: Vodafone UK's Tariff Structures, June 2001

	Leisure 200	Leisure 500	Voda 20	Voda 60	Voda 150	Voda 350	Voda 700	Voda 1000	Voda 2000	Voda 5000	Voda 10000	Voda Business
Connection (£)	35	35	35	35	35	35	35	35	41.13	41.13	41.13	30
Monthly line rental (£)	14.99	14.99	12.99	17.5	25	40	70	95	180	450	900	14
Monthly mins included	200	500	20	60	150	350	700	1000	2000	5000	10000	0
National — peak (pence per minute)	15	34	15	15	10	10	10	10	10	10	10	10
National — off peak (pence per minute)	2	2	5	5	5	5	5	5	5	5	5	5
On net — peak (pence per minute)	10	10	10	10	10	10	10	10	8.2	7	7	10
On net — off peak (pence per minute)	5	5	5	5	5	5	5	5	5	5	5	5
Off net — peak (pence per minute)	50	50	50	50	50	35	35	35	35	35	35	35
Off net — off peak (pence per minute)	30	30	30	30	30	25	25	25	25	25	25	25
Voicemail — peak (pence per minute)	10	10	10	10	10	10	10	10	10	10	10	10
Voicemail — off peak (pence per	5	5	5	5	5	5	5	5	5	5	5	5

Source: www.vodafone.co.uk

Many retail-pricing packages involve "two-part tariffs", with the subscriber paying a fixed fee and a per-minute charge.[40] In the above table, the Voda Business price plan involves a £30 (€48.82) connection fee and a monthly line rental of £14 (€22.78) per month before the customer has made any call minutes.

A consequence of two-part tariffs is that the average price of a call minute will depend on the total number of call minutes the subscriber makes. Without an assumption about the number of call minutes a subscriber makes, there is no way of knowing the average price of a call minute, beyond dividing total revenues by total call minutes. It is not possible at a disaggregated level to compare the average cost of call minutes with the average price of call minutes without knowing the number of call minutes the customer makes.

A cost model might have non-traffic-related costs and traffic-related costs. The latter could be compared with the price per minute that the MNO levies on subscribers. Except that there will typically be a variety of tariff structures available, with customers able to choose between packages that have a high fixed fee, low per minute charge or a low fixed fee, high per minute charge. There will not be a single price per minute faced by all customers. Nor need this be desirable. To continue operating, a MNO needs to recover all network costs, both traffic and non-traffic related. Setting per-minute prices equal to some measure of the cost of a call minute for all customers, and then setting the fixed charge equal to the non-traffic related costs divided by the number of subscribers, might result in a fixed charge that was in excess of what some low volumes customers are willing to pay. If it is not possible to discriminate directly between customers when setting the fixed charge, a solution might be to offer a variety of price structures and let the customers self select: high-volume users will choose a price structure with high fixed

40 Many MNOs offer handset subsidies. These can be thought of as negative fixed fees.

fees and low per-minute charges (contract customers), while low-volume users will choose a higher per-minute charge to and lower fixed fees (pay-as-you-go customers).

A second problem when attempting to compare prices of individual services with the costs of those services is that typically consumers will be offered tariff structures that entail a bundle of goods with pre-specified prices. Where one plan might have a low price for off-peak calls to fixed networks, another plan with higher prices for such calls might have lower prices for peak-time calls. Customers will select the plan that best suits their needs. But if two sets of customers are paying different peak rates (and different off-peak rates) it cannot be the case that they are both paying cost-reflective prices on a call-by-call basis. Such bundling is another way that an MNO may be able to recover non-traffic related costs: customers are offered attractive rates on those call types that are most important to them, whilst paying a price above the traffic-related cost for other types of calls. The rationale for having some prices differ to the incremental cost of the service applies also to the prices for roaming and call termination.

There is a cost-based rationale for differences in the prices for peak and off-peak calls. The former influence the size of the network that has to be built. Some elements in the network would not be necessary to serve off-peak traffic. By setting differential prices according to demand at different times of the day, the MNO can encourage callers to make more calls in off-peak periods, thus reducing the size of the network that needs to be built and therefore the total cost of the network. If price caps are to be set on call termination rates, there may be merit in allowing some flexibility by allowing the MNOs to set different rates at different times of the day, as is the case in Italy. More generally, offering different tariff bundles will allow an MNO to target different customer groups. If the MNO gets the desired mix of right, this may be to the benefit to all subscribers since network utilisation will be relatively high throughout the network at all times.

7.2.1 Price baskets and call origination

There have been a variety of approaches taken to overcome the sorts of problems discussed above with identifying a price for services. For example, INTUG's 1999 study looking at roaming prices calculated the price under a business subscription of an international roaming call to the home country in peak hours with a duration of 2 minutes 15 seconds. Oftel in the UK calculate the price for various bundles of calls using the cheapest plan — see below.

For the reasons outlined above, one common approach to facilitate price comparisons is to start by classifying mobile customers into particular categories according to usage patterns and then apply operator tariffs to these profiles to develop price baskets. It is common to differentiate between high- and low-usage customers, often by distinguishing between personal and business users. The price baskets may have to be re-modelled to reflect developments in the tariff and usage patterns of the market.

It is possible using this approach to compare prices across countries and over time, although there are limitations. The structure of mobile phone services and discounts are complicated. For example, MNOs offer different peak/ off-peak timings, free calls, per second billing and "friends and family" type schemes as ways of attracting customers. Many of these factors are not taken into account in a price basket index. Moreover, the basket of calls for which a price is calculated will have implications for which package is cheapest.

The OECD presented one set of price data in their report on Cellular Mobile Pricing and Trends in May 1999. In that report they referred to price baskets developed by Teligen, which made the basic distinction between a personal and business mobile phone user. The personal mobile basket excluded international calls and comprised 568 calls involving both a fixed and usage component. The Business Mobile basket presented by the OECD in May 1999 included 1,169 calls, incorporating volume discounts and international calls. While there is a need to compare prices for similar baskets if wishing to make cross-country comparisons, the basket chosen may not be representative of "typical" users within a given country.

Oftel in the UK have adopted another model to track price changes over time and reflect the experience of different types of customer. They do not make the distinction between a personal and business user but instead user profiles were identified from responses to a customer survey carried out in 1999. The model distinguishes between contract and pre-pay customers, and contract customers are further subdivided into those who pay subscription charges on a monthly basis and those who subscribe to an advanced payment scheme since the system of pricing used within these two segments differs considerably in the UK.

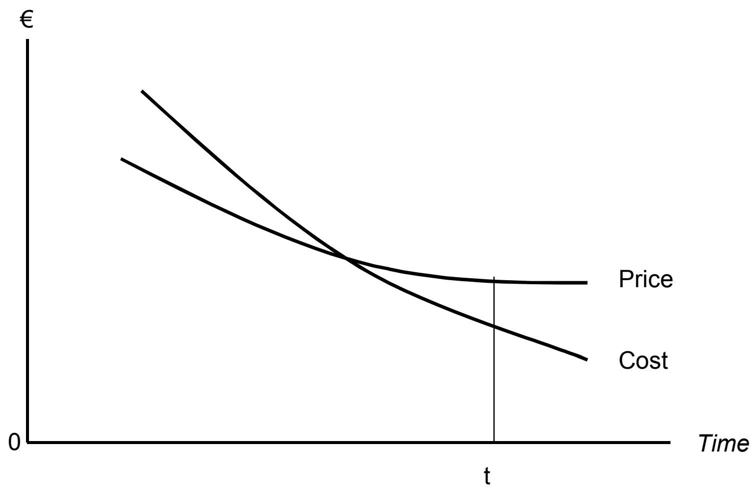
The baskets of mobile services reflect the typical mobile use for each category of customer, i.e. the proportion of calls made at peak, off- peak and weekend rates, the proportion of calls made to fixed and mobile networks and the use of other services such as SMS. The price for each basket of services is calculated for each MNO, optimising over the various packages offered within the pre-pay, subscription and contract categories.

Using a weighted average for all prices suggests that prices had fallen by 16 per cent between June 1999 and June 2000. But this masks very big differences between the baskets. The pre-pay basket had fallen by a third, whereas advance and monthly contract prices had fallen by 2 and 12 per cent respectively. This again illustrates the problem with expecting too close a relationship between prices and costs, even in an allegedly competitive market. At first glance, since all three baskets entail call origination services and, arguably, an access charge, the trend in costs of providing the three baskets should be broadly similar. Yet the price trends are very different.

7.3 Dynamic Profile of Prices

Even after a methodology has been developed to arrive at a single price paid by a subscriber for the basket of mobile services, there are still problems with attempting to compare price with cost. Focussing on the prices and costs at a single date may be misleading, since it ignores dynamic considerations. For example, an MNO may decide to price below cost initially to attract a large subscriber base so that it can realise economies of scale. At some point the MNO will need to price above costs to recoup the losses incurred while it was seeking to attract subscribers. As Figure 7.2 shows, it is possible that the prices will have fallen over time, but at a slower rate than costs. Consumers have consequently benefited from the MNOs strategy of attracting subscribers and realising cost savings, even though they end up paying prices above cost.

Figure 7.2: Price and Costs over Time



At time t, what cost should price be compared with? The cost incurred at time t, or the cost that would have been incurred had the network operator set cost-based prices at earlier dates. In the latter case, the prices at earlier dates would have been higher, which would likely have reduced the number of subscribers joining the network, the network size would have been smaller, resulting in higher costs as a consequence.

This line of reasoning is not too dissimilar to the rationale for permitting patents in other industries. For example, in pharmaceuticals companies can patent new drugs and act as monopoly sellers for 20 years. This can result in prices significantly above the level that would arise in a competitive market with generic manufacturers competing. But if competition were permitted instantly, i.e. there was no patent protection, then the incentive to undertake research and development in the first place would not be present: the drug would never have been developed by the private sector. Similarly, to insist on cost-based prices would limit the incentive for MNOs to roll out new technologies. It would cap the upside risk while requiring the MNOs to incur the risk that the investment is unsuccessful.

7.4 Cross-Country Price Comparisons

An initial investigation into mobile prices might compare prices for similar services across countries. How much can be learnt from such an approach?

7.4.1 Comparing the price of one service

Mobile networks are characterised by substantial joint costs between the services users purchase. Different services will bear different proportions of these joint costs. In both monopolistic and competitive markets, the prices charged will depend on the demand firms face for the various services they provide. Firms will seek to recover a higher proportion of joint costs on services for which they face demand that is less sensitive to price. An implication of this is that if demand conditions vary between countries — as they almost certainly do — this will result in different prices for the same services. This applies even if the networks are identical in every technical respect.

Therefore, observing different call termination rates in different countries (countries that may be deemed very similar in network design and size) does not imply that the mobile sector in one country is more competitive than the mobile sector in another. Nor does observing relatively low termination rates in one country imply that networks in this country are relatively efficient. If joint costs are recovered across different services in different ratios then looking at one service alone does not reveal much about competitive pressure or network efficiency.

7.4.2 Comparing the price of multiple services

If we compare very similar networks, we might expect countries that have high termination rates to have low charges for other services, such as origination and subscription. But as the number of services offered grows, and the range of different tariff options increases, an exercise of comparing bundles of prices between networks will prove more difficult. Nevertheless this is a more correct approach (given that joint costs are large) than comparing the price of one service between networks.

7.4.3 Prices across time

One more factor should be taken into account when comparing prices between countries. Pricing is not just a decision across different services but also over time. In a dynamic setting, we would not expect networks in different countries to take exactly the same path of development. In particular, again depending on demand conditions, networks in some countries may price all services very low to quickly gain a subscriber base and exploit economies of scale later on. Meanwhile networks in other countries may more slowly gain subscribers, but keep prices more reflective of annualised costs.

However, a prolonged period of time where prices for all services are above those on a comparable network may indicate absence of competitive pressures and/ or inefficiency, and suggest the need for further investigation.

7.4.4 Best-practice price caps

Best-practice price caps set regulated prices in one country based on prices levied for similar services in other countries. For example the termination rate in one country could be set based on the average of the three cheapest termination rates for a GSM networks in the EU, where the three cheapest rates are thought to reflect best practice for termination. [41] The justification for such an approach would be that this price better reflects the efficient charge for a particular service or is more appropriate on equity grounds than what is currently charged.

---

[41] Note that this "best practice" between countries is not just about efficiency and competitive pricing, but involves the outcome of decisions over the allocation of joint costs. This does not conform to what is normally thought of in "best practice" which tends to indicate efficient or competitive pricing alone.

An important consideration is whether the best-practice price chosen is an unregulated or regulated one.

7.4.4.1 Benchmarking against an unregulated price

As discussed above, if different services include significant joint costs then looking at best-practice prices in one service alone is erroneous. Neither the extent of competition in the mobile sector nor the efficiency of the network can be understood from one price alone. A price cap on one service would not just reflect the costs of that service, but how — given demand conditions — the operator has chosen to recover joint costs.

Unless the regulator is very confident that both the cost and demand structures are very similar in the two countries, a best-practice price cap on one service would be inappropriate.

7.4.4.2 Benchmarking against a regulated price

An alternative to setting a price-cap on unregulated best-practice elsewhere is to set it with reference to a regulated price. For example call termination rates could be based on what a regulator in a different country has found to be cost-oriented termination charges.

An obvious benefit of this approach is that it reduces replication of work across NRAs, saving the cost of examining accounting data and/ or building a cost model. Set against this is the fact that if networks differ between countries, so will cost-orientated prices. Only if there is a clear case that cost structure, consumer demand characteristics and policy maker's objectives are very similar between two countries would it be appropriate to take the short-cut of using a regulated price from one to inform the best-practice price for another country.

7.4.5 Revenue comparisons

In discussing the merits of cross-country comparisons, the importance of looking at a bundle of prices rather than just one price was emphasised. At the highest level, the full bundle of prices multiplied by the relevant output volumes gives the firm's revenue.

If two networks are thought very similar in cost structure (including subscriber density and traffic volumes) then comparing total revenue might be useful. This approach escapes the problem of how different operators choose to allocate joint costs because no matter where these are recovered, they will show up under total revenue.

If one network shows persistently higher revenues than another, while delivering similar service volumes and qualities, this may indicate a less competitive mobile sector. (This approach cannot inform on how competitive individual markets are within the mobile sector, or signal an appropriate price or revenue stream for one service alone.) The caveat is that different networks and mobile sectors may develop in different ways, exhibiting different price paths over time. Looking at a snapshot of prices at one point in time is likely to be misleading.

Where networks differ between countries, in both levels of output and cost structure, it would be necessary to make adjustments before revenues can be compared. However, there is little chance of data existing that would allow estimation of cost drivers with the necessary degree of accuracy, on account of the large differences between networks in different countries and the rapid developments in the mobile sector over time. Once cost structures in different countries become important to the analysis, it seems far better to seek to understand the nature of the differences. This involves examining detailed accounting data and/or developing a cost model.

7.5 Implications for Developing Regulatory Framework

Under the new regulatory framework for telecommunication, national regulatory authorities (NRAs) who identify an MNO as having a dominant position will need to assess whether the position is being abused. The NRA will also need to understand the likely effects of different policies.

Perhaps the main conclusion coming from this section is that caution should be attached to placing too much emphasis on the relationship between costs and prices for any particular services. Prices are determined by both demand and supply conditions in a competitive market. When there are sizeable joint costs, prices will have to be above incremental costs for some goods. In a competitive market the mark-up may not be an equi-proportionate one. Nor do equi-proportionate mark-ups realise the goal of allocative efficiency, unless demand elasticities for all the services are the same. It is more important that a regulator considers the demand elasticities for the various services in a mobile setting than in a fixed setting because the proportion of joint costs is larger.

8 ROAMING

Summary

- In the wholesale market, the cost differences for domestic and roamed call origination are small, assuming that call origination is defined to include coverage.
- There are potentially larger cost differences for domestic and roamed call termination. But these cost differences alone are unlikely to explain differences in prices.
- For retail prices, the incremental and stand-alone costs may be very different because many of the costs for billing and administration are shared with domestic services.

8.1 Charging Arrangements

8.1.1 Wholesale

8.1.1.1 Normal Network Tariff

Prior to 1998 wholesale roaming charges were based on the Normal Network Tariff (NNT). The NNT was the tariff paid for a given service by the majority of subscribers of the network operator.

Initially outgoing roamed calls were set at the NNT for mobile-originated calls. Later, MNOs started to place a VPLMN (Visited Public Land Mobile Network) multiplier, which under the Standard International Roaming Agreement (STIRA), was capped at 15 per cent. The cap, introduced in 1995, was supposed to reflect subscription charges that would otherwise not have been reflected in the wholesale roaming charges for outgoing calls. Most network operators set their mark-up at 15 per cent.

The NNT was fixed for a period of time so changes, typically falls, in retail tariffs did not result in immediate reductions in wholesale roaming charges for outgoing calls. Moreover, operators could switch from a business-user to a residential-user tariff as the basis for the NNT, which typically increased the NNT.

The “calling party pays principle” in the EEA means that network operators did not (and still do not) charge their customers for terminating mobile calls. Consequently the NNT for incoming calls was zero: MNOs generally could not levy a wholesale charge for incoming roamed calls

8.1.1.2 Inter-Operator Tariff

Since 1998/9 a new regime has been in force for wholesale roaming charges following a fundamental revision by the GSM Association. The Inter-Operator Tariff (IOT) is the tariff a visited network operator levies on the home network operator for the use of the visited network. An IOT lasts at least six months and changes usually require a minimum of 60 days notice. However, unscheduled changes can occur at any time: to reduce the IOT; to cope with regulatory changes; to respond to changes in network interconnection costs; and when new services are introduced.

The IOT for outgoing roamed calls will typically depend on: destination (domestic or international, and possibly whether the call will terminate on a mobile or fixed network); the time of day (peak and off-peak); the length of call, measured by units that may be considerably greater than per second; and a fixed fee for call set-up. The IOT for international calls will typically involve some form of zonal pricing, with the same rate applying for more than one country; some operators even apply the same IOT for international roamed calls in the EEA and domestic roamed calls in the same time zone. The definition of peak and off-peak differs by operator, since operators usually differentiate between these periods according to demand characteristics, and these may differ by network. To simplify marketing and minimise transaction costs some operators apply a uniform tariff for peak and off-peak, in some cases by applying the peak rate at all times.

There may be some discounts given under the new IOT regime. However, the extent of such discounts is unclear.

The new IOT arrangement means that operators can levy a wholesale charge on incoming roamed calls above their standard termination rate; such charges no longer have to be related to call termination charges levied to call the network's own customers. To date operators have not taken advantage of this change.

For roamed calls, the MSC of the visited network passes on call details to the network operator's billing system. This system prices calls according to the IOT. The priced call is saved in a Transfer Account Procedure (TAP) file. TAP is the GSM Association's standard format for Call Record Data between roaming partners (CRD). The TAP file is then passed to a clearing house, typically via an EDI link. The clearing house will send a validation report to the visited network requesting a correct transmission if there are errors. MNOs typically outsource these activities to permit realisation of economies of scale.

TAP3 is the latest version of the standard. It allows discounting at the invoice level. Unlike previous standards, TAP3 permits instant billing, rather than a delay of between 7 and 10 days between the call being made and the subscriber's network operator becoming aware of the call. The previous arrangement meant that MNOs were reluctant to offer roaming to their pre-pay customers since it was difficult for the MNO to ensure that the subscriber did not exceed their credit limit.

8.1.2 Retail

The calling party pays principle applies for outgoing roamed calls, just as it does for non-roamed calls. Most retail customers pay the same rate for roaming calls irrespective of the pricing plan they are on.

Network operators are free to price retail roaming services as they choose; there is no restriction in the GSM Association's Charging and Accounting Materials. Yet typically the home network operator adds a handling charge between 10 and 40 per cent to the wholesale roamed price levied by the operator of the visited network. The visited network operator's IOT determines the structure of the home network operator's retail roaming tariffs. Tied or independent service providers will typically share the handling charge.

Recently operators have started to offer averaged retail roaming prices. These may involve a single retail roaming price for roaming services in a particular country that does not differentiate between alternative visited networks. In other cases, mobile operators offer a flat rate for calls to the home country from any EU or EEA countries.

An exception to the calling party pays principle in European mobile telephony is for terminating roamed calls. In such instances the calling party pays the standard rate that applies to calling the called party's home network. The called party pays for re-routing the call from the subscriber's normal network.

8.2 Network Costs of Roaming

For each of these call categories, location updating must have been successful before a call can be made or received by the roaming subscriber. Appendix A2 provides some more details on the procedures for the origination and termination of calls by a network providing roaming services to a foreign operator's subscriber. This section focuses on the wholesale cost differences.

8.2.1 Originating a call when roaming

The information necessary to enable the roaming subscriber to make and receive calls has been copied from his/her home network HLR to the visited network VLR in location updating. Therefore, when the roaming subscriber makes a call, the call origination procedure is identical to the case where the visited network operator is originating the call from one of its own mobile subscribers. The costs of originating the call should therefore be the same as if a corresponding call was made by one of the visited network's own subscribers. The corresponding calls that we might compare with different types of roamed calls are identified below.

- *In-country calls.* There is no difference in terms of call routing between a network originating an in country call from a roaming subscriber and originating a domestic call from one of its own subscribers.

- *Calls home.* A visited mobile operator originating a call home by a subscriber roaming on its network will undertake the same procedure as if it was originating an international call from one of its own subscribers to a fixed or mobile network in the roaming subscriber's home country.

- *Country-to-country calls.* Originating a country-to-country call from a roaming subscriber is identical to the visited network originating an international call (to that country) from one of its own subscribers.

8.2.2 Receiving a call when roaming

The total cost of a call to a roaming subscriber will depend on the location of the two parties involved in the call and the route the call takes between these.

There are three networks involved in the call (ignoring any network that undertakes transit on behalf of one of these). The home (mobile) network of the roaming subscriber, the visited (mobile) network that the subscriber is roaming on and the originating (mobile or fixed) network that the originating caller is connected to.

As before, the three types of calls that might be considered are:

- *In-country calls.* Although there is no international dimension to the conversation that takes place, an in-country call is international in terms of call routing. The originating and visited networks are in the same country, but the call is international in nature because the call is routed between these two networks via the home network (which is in a different country).

- *Calls home.* When a call to a roaming subscriber is originated in the subscriber's home country the call only has one international leg, that from the subscriber's home network to the foreign network the subscriber is roaming on.

- *Country-to-country calls.* A country-to-country call will have two international legs; all three networks involved in the call are in different countries.

The table below summarises, for these three call types, the networks that would be in the same country and those that would be in different countries.

Table 8.1: Comparing Roamed with Non-Roamed Calls

	Networks in the same country	Networks in different countries
In-country call	originating & visited networks	home network
Home country call	originating & home networks	visiting networks
Country-to-country call	-	originating, home & visited networks

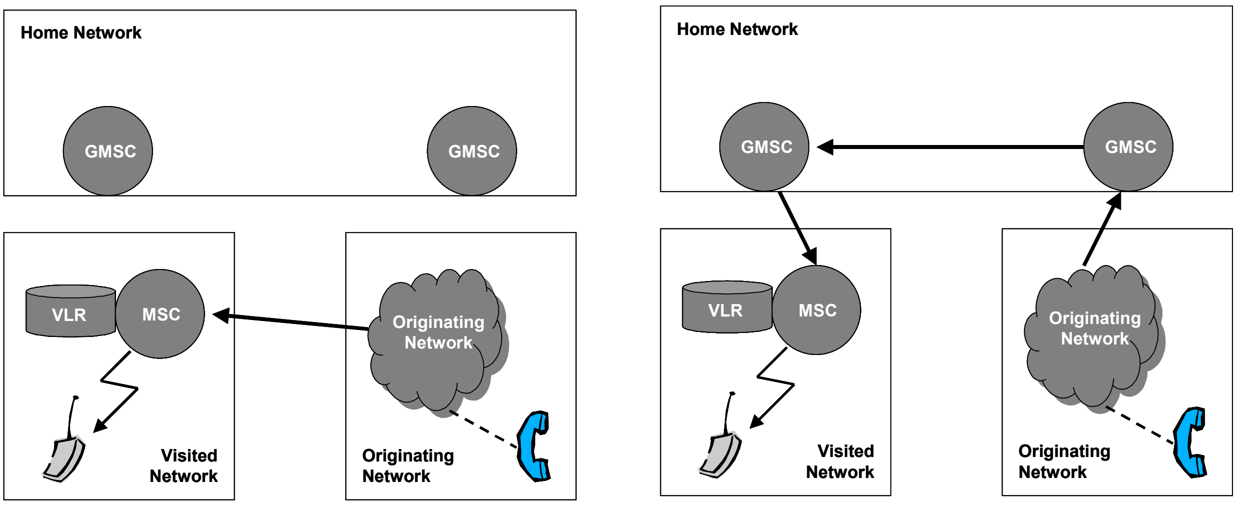
Only the originating and visited networks are essential for the voice transit (i.e. excluding set-up and signalling) of the call. But in many cases the call is conveyed across the home network.

There are two alternative ways that the call can be terminated. In the first, once the home network HLR has located the visited network VLR and retrieved the MSRN, the call is conveyed from the originating network to the home network, across the home network, and then from the home network to the visited network where the call is terminated on the roamers handset.

Alternatively the call could travel directly from the originating network to the visited network, bypassing the home network altogether.[42] Which of these routes the call takes will depend on both the technology available and the agreements in place between networks.

The diagrams below ignore the signalling involved in call set-up and concentrate on the alternative routes the call can take between the two calling parties.

Diagram 8.1: Alternative call transit routes for call termination



Bypassing the home network *Call transit via the home network*

This bypass is likely to be of greatest advantage for in-country calls because the voice transit will then be entirely domestic. Even for a country-to-country call the saving due to more efficient routing could be large. For example, consider a call from France to a Japanese GSM subscriber roaming in Germany. If the traffic travels directly from the French to German network, costs for conveying the call could be significantly lower than if the call had to go via Japan.

Generally it will be inefficient if calls are routed through the home network. If the home network can be bypassed, clearly the costs associated with a call to a roaming mobile customer will be lower than if such bypass is not possible. Even if it is technically feasible, an MNO may not use the technology, possibly because a rival has proprietary rights to it. In a regulatory context, forward-looking cost models will need to make an assumption about whether bypass is possible. Once the home network adopts CAMEL Phase II technology such bypass will be feasible.

8.2.3 Receiving party pays — the costs

An exception to the general principle that the calling party pays applies when a roaming subscriber receives a call. The rationale for this is that the calling party should have certainty about the price that will apply when calling a mobile number, since the calling party will not know whether the receiving party is roaming or not. It is instructive to compare the costs the visited

---

[42] Note that for calls that originate from a roaming subscriber, the subscriber's home network is not involved in the call, which passes direct from the visited network to the receiving party's network.

network incurs when a terminating a call for someone roaming with the costs incurred by a home network terminating a call for a subscriber, i.e. someone not roaming.[43] This is the *termination cost differential* due to the fact that the subscriber is roaming — what the receiving party would pay when roaming if the charge was cost-based.

Regardless of which of the three types of roaming calls are to be terminated on the visited network, the termination cost to the home network should be very similar.[44] This is because all three involve the (international) information exchange between HLR and the visited network VLR, the (international) call transit from the home network to the visited country,[45] and the final conveyance of the call across the visited network to the roaming subscriber.

If the home network and visited network had identical cost structures, this cost differential would be approximately the cost of international transit of the voice traffic from the home network to the visited network. Where the networks differ in terms of the costs of terminating calls to their own subscribers, the cost differential will take account of both the international transit and the differences in network costs between the operators.

Suppose the visited network has higher termination costs to its own subscribers than the home network, perhaps because of geographical features or subscriber density.[46] In this case the additional cost, because the subscriber is roaming, of terminating the call on the visited network rather than the home network is more than the cost of international transit between the two networks. In contrast, if the visited network has lower termination costs than the home network, the differential between terminating roaming and non-roaming traffic will be less than the international transit cost. Note that it is possible that the cost of terminating a call to a subscriber roaming abroad could be less than the cost of terminating a call to a subscriber on the home network. This may be the case if the visited network termination costs are low compared to the home network, and the international transit costs are sufficiently small.

8.2.4 Incremental cost of wholesale roaming services

As in Section 4's discussion, to calculate the incremental cost of wholesale roaming services requires first a definition of the increment or increments. One candidate would be to compare the costs of a network that offered roaming services, including a specified number of call origination and call termination minutes against the costs of a network that did not. An alternative would be to look at smaller increments. For example, one increment could measure the incremental costs of being able to offer roaming services, without including any call minutes. And then to consider the incremental costs of the call origination and the call termination minutes.

This latter approach might more sensibly just include the costs of roamed call origination minutes in the same increment as domestic call origination minutes. Once a roaming subscriber's location has been updated, call origination has the same cost structure regardless of whether the user is

43 Total costs refer to the costs incurred by both the home and visited network, ignoring any wholesale roaming prices that the two parties levy on one another.

44 Small variations may come about where different calls enter the home network at different GMSCs.

45 This assumes that the technologies are not in place to permit call transit to bypass the home network.

46 Section 2.2 of the report discusses the main cost drivers of a mobile network.

roaming or is a home subscriber. This assumes that coverage is included in the definition of call origination for domestic users. If it is not, then there might be a distinction to make between the two sets of services, since it might be argued that when roaming, the call origination does include a charge for coverage whereas domestic subscribers are paying for the service of coverage separately, perhaps as a subscription charge.

Whether coverage is a separate service or not also applies when considering whether roamed call termination minutes should be included in the same increment as domestic call termination minutes. However, for call termination, there is no need for interaction with the HLR in the visited network, so the cost of terminating calls from GMSC to a roaming subscriber's handset is slightly less than for terminating a home subscriber's calls. If this difference is deemed important there is a case for separating the services of termination to home users and termination to roaming users. Different per minute costs can then be derived for the two, either through cost allocation within a voice traffic increment, or by defining these services as separate increments.

There are other incremental costs associated with a network offering roaming services to other networks, beyond those associated with offering call origination and call termination minutes. There are different subscriber costs in providing services to an international roaming user rather than a home subscriber. These can be included under the increment of roaming subscribers. Since roaming subscribers' have their SIM card and handset provided by their home network, the cost of this increment will be the incremental cost of location updates for N roaming subscribers. This will be relatively small, just consisting of extra VLR database capacity and the extra network capacity required for the information exchange between VLR and home network HLR. A network also incur the costs of a roaming platform, costs associated with reaching a bilateral agreement, costs paid to clearing houses, and costs associated with testing the system to be sure that the system works. These are incremental costs incurred to offer the roaming service, although they are not incremental to any particular customer.

There is even a question of who should pay for these incremental costs. The importance of defining who is purchasing what service reappears. Are the costs associated with offering roaming services ones that subscribers to a network should pay, since they are costs necessary to offer the subscriber the ability to roam should the subscriber go abroad? Or are the costs ones that roaming customers should bear? If the latter, then the costs of call origination and call termination might have to be higher than for equivalent domestic calls, assuming no fixed charge for the facility to roam applies. As before, efficient pricing principles should determine the relative mark-ups. But there is a case for arguing that network competition in the domestic markets should extend to competition to offer the best deals on roaming. Therefore subscribers pay for the costs that their own network has incurred to allow them to roam abroad; the actual costs included in the wholesale charge should not include a charge for setting up roaming facilities since these are costs incurred providing a service to domestic customers.

8.3 Comparing Retail Roaming Prices and Costs

For many of the reasons discussed in Section 7, there are important caveats to be attached when seeking to compare the retail prices of roaming with the costs of those services. It is also important to determine which costs are to be compared with prices. It could be the costs the subscriber's own network incurs to offer the roaming services. These costs will include the wholesale charge that is levied on the network operator when the MNO's subscriber roams. Alternatively the focus could be on the actual network costs incurred to permit a subscriber to roam on visited networks, along with relevant retail costs incurred by the home network. This latter approach might be a better approach if the goal is to establish whether the prices charged for roaming are efficient from society's perspective.

But if the focus is on whether prices in the retail market are competitive, then the wholesale price charged to networks offering their customers roaming should be included in the costs calculated for the roaming service. If these wholesale charges are too high, that should be determined from a study of the wholesale market for roaming services.

Aside from the network costs of offering roaming services, there are also retail costs that need to be considered. To offer end-user customers a roaming service the firm will incur various administrative costs, including the costs of billing and advertising. Many of these costs would also be incurred if the firm did not offer roaming facilities. Customers would still receive bills if they could only make calls in their home country; the costs of advertisements to attract customers and to encourage them to use their phone will in many cases be difficult to attribute between roaming and domestic mobile services. Many of the retail costs are joint costs, joint to both the roaming and non-roaming services that subscribers purchase.

Consequently, just focussing on costs to explain the differences in prices would be mistaken. The appropriate mark-up over the incremental cost of a service will depend on demand conditions; even if the incremental costs of call origination are the same for home subscribers and roaming customers, there may be a case for different prices if there are different sensitivities of demand to price. This may have been true when roaming was confined mainly to contract customers, who were often business customers. But an alternative explanation for the reduced price sensitivity of roaming customers is a lack of transparency about the prices. If true, this would be a cause for regulatory concern and measures might be sought to increase transparency before the relative demand elasticities of roaming and non-roaming customers were compared.

Some retail costs that can be attributed to the roaming services include costs associated with fraud or default risk. However, the costs are not attributable to either call origination or call termination. Again there is a case for a mark-up that should be determined with reference to demand conditions. The problems apply more to prepaid customers. Some of the problems may be lessening with the introduction of CAMEL Phase II. This will permit prepaid roaming without the risk of customers spending more than they have because the time delay under the conventional TAP arrangements will be removed.

9 REGULATING MOBILE COSTS AND PRICES

9.1 Introduction

This report has set out the cost drivers and characteristics of mobile networks, and explained how far, and in what ways, the prices of different mobile services might be expected to reflect costs.

In each section, it has sought to show the consequences for the practice of regulation. This brief final section brings the main points together.

Analysis of whether to regulate the mobile sector in a particular member state lies wholly outside the scope of this report. We simply note that 24 MNOs in the EU have been designated as having SMP in the mobile market. Of these, 12 MNOs in 8 member states have also been notified with SMP in the market for interconnection. NRAs have imposed a range of obligations on those MNOs: at the time of our survey, price controls on mobile termination charges applied in 8 Member States.

The cost structures of mobile networks, and the relationship of costs to prices, have important implications for the choice and specification of regulatory obligations.

9.2 How Mobile is Different From Fixed

The trajectory of regulation in fixed telephone networks in the EU is well-known. The key documents are listed in Section 2.

In a mobile network, the customer has the right to access the network at any point, not just at one point. The access points are all shared. The technical differences in mobile telephony are not trivial, but have fundamental effects on the allocation of costs between services.

The analysis of coverage is central to the difference between fixed and mobile networks. We prefer to define coverage as the capability or option to make a single call from any point of the network at a point in time. All additional capacity is then due to conveyance. But however coverage is defined, the element of joint costs will be larger as a proportion of total costs than in fixed networks. So, therefore, will the difference between incremental and stand-alone costs of a particular service.

Our analysis shows that a wide range of prices for a particular service may be consistent with cost-oriented pricing.

The other important differences relate to the different kinds of markets. In most member states, there are several competing MNOs. Technological change is rapid, and economic depreciation of assets is correspondingly rapid too. Companies can compete by offering different sets of services and different combinations of price and quality.

This makes the choice of regulatory objectives sharper and more important in regulating mobile telephony. Should the NRA's aim be to increase the level of competition in mobile services wherever possible? Is the NRA concerned about equity issues between different purchasers of mobile services? Does the NRA have strong concerns about allocative efficiency at a single point in time or are these dominated by a desire for innovation and competition for the development of new services over the longer term?

Competition and rapid change also make problematic the idea of estimating the costs of a notional efficient operator. Even where there is a single efficient model at time t, it will have moved on by the time a regulator has pinned it down at time t+1.

9.3 Weighing up Different Regulatory Strategies

Suppose an NRA decides that regulation is needed in the mobile sector. What are the implications of our analysis of the cost structures of mobile costs for different regulatory strategies?

Under the new Regulatory Framework, regulators will have more discretion when deciding on what regulatory measures to adopt. To avoid over-regulation, transparency, non-discrimination, accounting separation, access and price controls, including cost orientation will represent a maximum set of obligations that can be applied on undertakings with SMP.[47] In exceptional circumstances, NRAs will also have the power to impose obligations on all market players.

NRAs are obliged to ensure that measures adopted are effective, justified and proportionate. A measure is proportionate if a reasonable relationship exists between the regulatory action and what the regulator seeks to achieve, and the measure does not go beyond what is necessary to achieve the end result.

The intention is that the use of price controls should be to promote efficiency and sustainable competition and to maximise consumer benefits. For reasons discussed elsewhere in this report, a regulator will need to identify priorities before a decision can be taken on the need for and the level of any price control. As Section 2.5 discusses, there are potential conflicts between realising allocative efficiency and productive efficiency. The EC's Draft Directive on Access and Interconnection warns that price controls should not negatively affect competition in the long run or reduce the incentive to invest in alternative facilities in the long run.[48] Equity considerations may further complicate matters. Controlling prices for one service may well affect the prices of other services: a price control may not benefit all consumers. The large pot of joint costs means that demand conditions will also have to be considered if consumer benefits are to be maximised. A focus solely on costs to set "cost-oriented" prices is unlikely to be effective, justified and proportionate.

9.3.1 Strategies that make sense

Regulatory strategies that make sense will be those that:

- specify objectives;

---

47 See paragraph 9, Com (2001) 369.
48 Article 13(1), Com (2001) 369.

- define services carefully — we suggest including coverage in services such as call origination and call termination;
- look at effectiveness of competition in mobile sector as a whole and do not without very good reason reduce that competition or the incentives for further innovation; and
- allow in the regulation of prices for there being substantial joint costs.

9.3.2 Strategies that make no sense

Regulatory strategies that do not meet these conditions may do harm than good. Obvious traps include:

- using some notion of total network cost divided by total number of minutes as a basis for making recommendations on prices of individual services — such estimates may bear no relation to the cost of providing any particular service;
- comparing prices for individual services with prices in other countries without acknowledging that competition may be working differently;
- examining the price of one service in relation to costs without looking at the prices of other services; and
- imposing arbitrary rules for mark-ups on cost when setting prices for individual services, thus preventing price competition.

9.4 Non-discrimination

On-net calls use the same network operation as call termination. Both require a call to be terminated to one of the network's subscribers. Although an on-net call does not come in from another network, it will be received by an MSC (acting the same way as a GMSC) and the same requirement to find and notify the called subscriber exists as for off-net calls. Therefore, firms following the non-discrimination principle, will not price on-net calls below the wholesale termination rate (for the relevant time of day).

Therefore regulators could check for non-discrimination by comparing on-net call charges with operator's termination rates. If the on-net call charges are below wholesale termination rates, prices should be realigned to satisfy the principle of non-discrimination. But prior to applying the principle, the regulator should consider how important non-discriminatory prices are relative to other objectives the regulator may have. Low charges for on-net calls are one of the ways that MNOs compete to attract customers, so regulatory intervention to stop non-discrimination would limit competition between MNOs. Furthermore, neither the competitive price, nor the price that would arise if the non-discrimination principle was applied to prevent on-net call charges being below wholesale termination rate, will necessarily correspond to the price of on-net calls that realises allocative efficiency. The efficient price for on-net calls will depend on the elasticities of demand for the various services at the industry level; the competitive price will reflect the elasticities of demand at the firm level; they may not be the same.

9.5 The Value of Regulatory Accounts

Regulatory accounts can help ensure that wholesale and, in some cases, retail tariffs offered by operators designated as having significant market power (SMP) are cost oriented, transparent and non-discriminatory.[49]

Although regulatory accounts are often an obligation imposed on operators with SMP, few operators produce a meaningful set of regulatory accounts. BT in the UK is often cited as an example of an operator producing useful accounts for their fixed network. The information provided includes the costs of some key network components, volume information and per minute charges. BT is also required to produce a statement of regulatory accounting principles, explanations of the attribution methods used, statements of the accounting policies used and the methodology used to calculate long-run incremental costs.

Regulatory accounts may have an important role to play in regulating mobile operators. They could, for example, be used to show the split between retail and wholesale costs and to monitor whether the wholesale business is discriminating between what they charge their own retail business and other access seekers. They could also be used as a way to monitor whether prices for some services (such as call termination) are above the stand-alone costs of the service.

Preparing a useful set of regulatory accounts will require significant investment by both NRAs and by operators. NRAs will need to determine how the accounts will need to be prepared. We strongly recommend that any accounts that are to be produced are capable of identifying stand-alone, incremental and joint costs. The Commission has indicated to us that it may also contribute recommendations for a common approach on the publication of regulatory accounts of SMP mobile operators with cost-oriented obligations. NRAs will have to define the separate businesses and services for which accounting information will need to be collected. They will also need to agree on the attribution methods, cost bases, and the accounting rules to use. All of these will require consultation with mobile operators and perhaps wider consultation with other interested parties.

Mobile operators will also incur costs when preparing regulatory accounts. Regulatory accounts do not reflect the way that operators are actually structured. They require different sets of numbers to be produced in addition to those required for statutory and management accounts.

In short, regulatory accounts have the potential to be a rich source of information for regulatory authorities and other operators, but do require significant effort to implement properly. Our understanding is that the European Commission has recently commissioned a study to assess the various practices and initiatives by National Regulatory Authorities with regard to cost accounting and accounting separation methods applied by notified operators. The findings of this review could be used by regulatory authorities when deciding whether, and how, to implement effective regulatory accounts.

---

[49] The Commission Recommendation of 8 April 1998 provides a common approach for accounting separation and cost accounting of fixed network SMP operators.

9.6 Building Cost Models

Regulatory authorities have a great deal of experience with developing cost models to estimate interconnection charges for fixed networks. These costs models can either be used to set charges, or to challenge the results from costing models developed by incumbent operators. These models typically take some time to build and then some time to reconcile. Typically between eight months to two years is needed before sensible results are obtained.

There may be a role for costing models in mobile as well. These could be used to estimate the costs of call termination and possibly to resolve disputes about the terms of indirect access or interconnection (e.g. national roaming). This may be particularly important when new 3G licence holders want to gain access to existing 2G networks.

But regulatory authorities should note that it takes time and there will be additional complications in mobile. First, it is more dynamic. Second, the definition of services and increments is more problematic. Third, the coverage issue means that rules about allocating joint costs are more important if trying to set prices.

The regulator should be transparent about what is being done. Because joint costs are important, they should be clearly identified. The regulator should explain the rule(s) used to allocate the joint costs, and what objective is achieved by following such a rule.

APPENDIX 1: AVERAGE COST PER MINUTE CALCULATIONS

DG Information Society requested, in the absence of detailed accounting information available from mobile operators, crude estimates of the wholesale cost per minute based on the limited data available in published accounts.

For reasons outlined in the report, regulatory intervention should not depend solely on estimates of average costs. Nor should the accuracy of the estimates presented here be relied upon to formulate policy responses.

A1.1 Wholesale vs. Retail Costs

In the published financial reports, mobile phone companies do not discriminate between the costs of incoming/outgoing calls on their network and the interconnection charges of incoming/outgoing calls to other networks. Hence, it is not possible to ascertain the real wholesale cost (on a per minute basis) using those reports. For the same reason, the retail cost is also very hard to determine; often, companies include in their cost of sales (which we use as a proxy for the wholesale cost) "…the cost of handsets and accessories…"[50] (e.g. Orange), as well as "…financial incentives to service providers and dealers for obtaining new customers…"[51] (e.g. Vodafone), which we would regard as components of the retail cost.

This makes direct comparisons between companies unreliable, and very sensitive to the particular assumptions made on the exact decomposition of those costs. The wholesale cost per minute which we calculate include the interconnection costs (and in Orange's case, the cost of handsets and accessories is also included) which inflates the estimate; against this, the calculations do not include a charge for the cost of capital employed.

Our analysis uses the Annual Reports published by five companies (Vodafone, Orange, Telecel Vodafone, Omnitel and Libertel Vodafone). Of these, only Libertel disaggregates their cost of sales into types of cost.

A1.1.1 Vodafone

The estimate of Vodafone Group's wholesale cost (on a per minute basis) includes "…payments to landline and mobile operators for delivering calls outside the Group's networks and for providing landline or microwave links…"[52] (i.e. interconnection costs) as well as "…depreciation of network infrastructure (…) and network operating costs."[53] (i.e. the real wholesale cost). Hence, we can only calculate an upper bound for the real wholesale cost. For the year ending on March 2001, the total cost of sales (including the interconnection costs) for the Vodafone Group was £5,338 million. The number of customers at the end of the year was 82,997,000. Vodafone does not report monthly average usage of the mobile phone, so we have used data from OFTEL related to the UK subsidiary, and applied that data to the group's results (see footnote 6). The

50 Orange Annual Report 2000, p. 95.
51 Vodafone Group plc Annual Report on Form 20-F for the year ended 31st March 2001, p. 45.
52 Vodafone Group plc Annual Report on Form 20-F for the year ended 31st March 2001, p. 45.
53 Vodafone Group plc Annual Report on Form 20-F for the year ended 31st March 2001, p. 45.

average monthly usage reported by OFTEL is 87.3 minutes,[54] and hence the upper bound of the wholesale cost is £0.061 per minute.[55]

A1.1.2 Telecel Vodafone

Telecel Vodafone is the second largest Portuguese operator, majority owned by the Vodafone Group. Again, in their 2000 Annual Report, they include the interconnection costs in the cost of sales (but not the cost of handsets and accessories, as in Orange's case). The total cost of sales was 221 million euros, and the total number of customers was 2,478,800, with a monthly average usage of their mobile phone of 135 minutes. This yields an upper bound for the wholesale cost of 0.055 euros per minute, lower than Orange's because the cost of the handsets and accessories is not included.

A1.1.3 Orange

Orange's cost of sales includes interconnection costs (as does Vodafone) as well as "…the cost of handsets and accessories…"[56], and is somewhat more unreliable than Vodafone's. For the year 2000, the total cost of sales was 5,358 million euros and the total number of customers was 31.2 million. The average monthly use of the mobile phone per customer was 147 minutes, which yields a maximum wholesale cost of 0.097 euros per minute.

A1.1.4 Omnitel

The Italian mobile phone operator Omnitel report a cost of sales (under the item "Industrial services and GSM network") of 836,968,601 euros in 1999. This figure includes "Telephone line rentals, access charges paid to other licensees for the use of their networks and roaming costs"[57] (i.e. it is equivalent to their wholesale cost plus their interconnection charges). Their total number of customers at the end of that year was 10,418,000, with an average monthly usage of their mobile phone of 135 minutes. This yields an upper bound on Omnitel's wholesale cost of 0.05 euros per minute.

A1.1.5 Libertel Vodafone

Libertel Vodafone, a Dutch mobile phone operator with a market share of 29% in the year ending on 31 March 2001, is the only operator we have found which disaggregates the cost of sales into types of cost. From their Annual Report 2000/2001, the cost of leasing lines and the depreciation of land, buildings and network assets (i.e. wholesale cost) was 94,988,000 euros. Using the customer base of 3,281,000 subscribers who, on average, use their mobile phone 111 minutes per month, we can calculate the per minute wholesale cost as being 0.022 euros per minute. In order to allow comparisons with the other operators we have mentioned, we could add to the wholesale cost the interconnection charges, which were 258,959,000 euros. This yields a cost

---

54 This is the average monthly usage in the 3rd quarter of 2000, as reported in OFTEL, Market Information Mobile Update: May 2001.

55 OFTEL reports an average monthly usage for Orange's subsidiary in the UK of 95.3 minutes, not too different from Vodafone's UK subsidiary. Using Orange's average monthly usage for the whole group (147 minutes) as a rough estimate for Vodafone's Group average monthly usage we obtain a maximum wholesale cost per minute of £0.036 (approx. 0.058 euros), not much different from Telecel's cost.

56 Orange Annual Report 2000, p. 95.

57 Omnitel Annual Report 1999, p. 142.

per minute of 0.081 euros (which can be compared to the upper bounds calculated for Vodafone, Telecel and Omnitel).

A1.2 Worked Example of LRAIC per minute

An alternative to dividing total costs at the network level by total minutes is to focus on the costs by network element.

An appropriate costing exercise would entail a careful examination of the topology of the network, the geographic and demographic features of the country under consideration and, last but not least, an accurate analysis of the demand side of the market. A bottom-up model would need to identify the total amount of equipment needed to serve demand.

A1.2.1 The methodology

As discussed in Section 3, the cost drivers of traffic and coverage area are cost drivers affect most of the costs of a mobile network. Potentially important cost categories include the cells (these include tower, antennae, and the site), TRXs, backhaul, BSCs, transmission in the backbone network, MSCs, VLRs, HLR, and network management equipment. There are both capital and operational costs that need to be calculated.

The approach adopted in this study does ***not*** work out the total amount of equipment needed to serve demand in peak time. Instead we develop a "micro" approach that looks at various network elements in isolation. The annual cost of each element is divided by the number of minutes the average element is assumed to serve during a year. Annual minutes are obtained through the conversion of Busy Hour Erlangs (this is a measure of Busy Hour traffic, which network elements are usually dimensioned to carry) into annual traffic.

The approach seeks to identify a cost per minute for each of the major network elements. Summing these gives an average cost per minute. This however does ***not*** provide the cost of a minute of origination/termination services. To get cost estimates for services, information would be needed on how different services use on average different network elements.

The approach described here implicitly uses equi-proportionate mark-ups to allocate the costs of coverage to different call minutes.[58] Section 7 discusses the usefulness of such an allocation rule in an industry where there are large joint costs. Identifying joint costs would require building a proper cost model. Pricing using equi-proportionate mark-ups may result in some potentially beneficial opportunities being foreclosed on account of the burden of joint costs allocated to them.

---

[58] This is also true of bottom-up models that calculate the total costs of the equipment needed and then divide by total minutes.

The results of using averages are least useful for those network components for which one of the following conditions applies:

- the main cost driver is not solely capacity; or
- the fixed costs are high.

In mobile networks, both these conditions apply when estimating the costs of the cells (tower, antennae and sites). Cell costs may be driven, and, in fact, in rural areas are, by the area covered rather than by the capacity carried. As with fixed networks, there are modularities in equipment size.

Among the different cost categories, it is the per minute cost of cells that is most sensitive to assumptions about population density. Given the overall impact of these costs on total costs, we distinguish, therefore, between two scenarios and estimate per minute costs of the following two different networks.

1 *Rural network*. Traffic over the network is such that the cost driver of each cell turns out to be coverage.

2 *Urban network*. Traffic over the network is such that the cost driver of each cell turns out to be traffic.

Our approach requires assumptions about the cost of the various network elements and about the annual number of call minutes that any element might serve under the two different scenarios.

Another caveat to be drawn refer to the role that data services take in this exercise. Data traffic as well as voice traffic makes use of all network elements which enter this exercise. The shared use of these elements will most probably lower the average per minute cost, given the relevance of the economies of scale in this industry. However, since data traffic still amounts to a low proportion of total amount of traffic and the inclusion of data services would require additional dimensioning assumptions, we do not take them into account.

In the next section we describe the relevant technical assumptions made for each network element under the two different scenarios. Then we describe the method used to convert BHE in annual minutes. The final section shows the results.

A1.2.2 The network elements

We approximate the number of Busy Hour Erlangs each of these elements serves (whether its cost is directly driven by capacity or not). In the following section we convert BHE in annual minutes and divide the unit cost of each element by this figure to get costs per minute.

Cells

The traffic carried in each sector will depend on demand considerations.

In the rural network, each cell is assumed to carry 4 BHE. Cells are under-utilised in this network, as a consequence of the fact that the number of sites in rural areas depends on the area to be covered rather than the traffic to be carried.

In urban network, each cell is assumed to carry 30 BHE. We assume that each cell covers three sectors, each serving approximately 10 BHE. Cell costs in urban areas are driven by the amount of traffic rather than coverage considerations.

TRXs

Transceivers are devices to be incorporated in each sector. Their number mainly depends on the volume of traffic to be carried.

We assume here that, given the amount of traffic carried and specific technical considerations, 1.3 TRX per sector would be needed in rural areas and 2.5 TRX per sector would be needed in urban areas. Under these assumptions, each TRX in rural areas would carry about 3 BHE and each TRX in urban areas would carry about 4 BHE.

Backhaul

An appropriate approach would estimate the costs associated with the cabling and trenching of the physical routes linking concentrator and switching centres. This would take into account appropriate sharing of logical routes. This approach is quite complicated and very data intensive. Moreover, the results are depend strongly on the country characteristics and topology.

To overcome this problem we assume that the transmission capabilities needed are provided through the use of leased lines, i.e. we assume that each cell is connected to its BSC through a dedicated circuit. This approach may lead to a proxy of transport costs if the prices are cost oriented. This is a requirement imposed on SMP operators by the European legislation; however, the general lack of wholesale offers (most European incumbents only have retail offers) means that we are likely to have overestimated the costs.

Costs of leased lines vary with the capacity carried and the length of the circuits. We assume that the average length of backhaul links is 20 Kilometres.

As to the capacity involved, a 2 Mbit/s leased line is here assumed to provide all the capacity needed to backhaul the cells with BSCs, in both rural and urban networks, even though the average traffic transported differs between the networks. A 2 Mbit/s link is here assumed to provide the minimum amount of available capacity.

BSC

BSCs are driven by the number of transceivers to be served. We assume that a Base Station Controller is needed for every 80 TRXs.

As explained above, number of TRXs is driven by capacity carried. As a result of these assumptions, therefore, BSCs serve approximately 250 and 300 BHE in rural and urban areas respectively.

BSC-MSC

As explained in the section on backhaul, an appropriate approach would design a transmission network based on the geographic/demographic characteristics of the country under consideration and cost it.

Once again we assume that this functionality is provided by dedicated circuits. For every BSC, a leased line is assumed to link it with an MSC. The length of these circuits is assumed to be 60 kilometres.

Given that BSC capacity is assumed to be driven by the number of TRXs, the capacity to be transported would not differ too much in the rural and urban networks. A 34 Mbit/s circuit is assumed to provide enough capacity for both the rural and the urban network.

MSC

A proper model would need to develop algorithms to determine the optimal total number of MSCs.

Here we use an alternative, less accurate, approach that consists of establishing an optimal average number of call attempts/number of minutes that each switch would serve. We assume that each MSC serves on average 300,000 call attempts in the Busy Hour. Busy Hour Call Attempts (BHCA) are then converted in annual minutes by assuming an average call duration of three minutes and then using the same conversion ratio used for BHE.

MSC ports

Ports are driven by the amount of traffic in the Busy Hour.

On average, each port in the MSC is assumed to serve 15 BHE.

MSC-MSC

We assume, once again, that transport in this part of the network is provided through dedicated circuits. The length of these circuits is assumed to be 150 kilometres.

It is difficult to estimate the capacity transported in this part of the network, not least because there is a lot of route sharing at this high level of the transmission hierarchy. Moreover, the dimensioning of this part of the network strongly depends on the total traffic the network is dimensioned for and, in this exercise, we do ***not*** take into account total traffic.

However, technical assumptions in the fixed framework would indicate that the capacity transported over each route of this layer of the network is between five and seven times the capacity transported over each route of the lower level.

Assuming that the average route in this layer of the network carries six times the amount of traffic carried in the average route in the lower level of the network, a 155 Mbit/s circuit is assumed to provide enough capacity to serve the approximately 1,500 BHE needed in the rural network, while a 34 Mbit/s circuit, as well as a155 Mbit/s circuit, is needed to serve the urban network's 1900 BHE.

HLR

Usually there is only one HLR per network. The HLR capability and hence cost is driven by the number of subscribers and therefore costs should be recovered on that basis.

However, since this exercise seeks to measure costs per minute of conversation (on an annual basis) of a hypothetical network, here we divide the HLR annualised costs by the total number of minutes carried over the network. As an approximation we divide by the number of minutes carried over an MSC.

Recap of the assumptions

The following table summarises the technical assumptions underlying the cost estimates of the network elements worked out in this section and their source.

Table A1: Technical assumptions underlying the cost estimates*

	Network A	Network B
No of sectors in macrocells	1	3
Average capacity per macrocell (BHE)	4	10
Average No. of TRXs per sector	1.3	2.5
Average length of backhaul (Kms)	20	20
Average capacity of backhaul (BHE)	4	30
Average capacity of backhaul (Mbit/s)	2	2
No. of TRXs served by the same BSC	80	80
Average length of the BSC – MSC (Kms)	60	60
Average capacity of the BSC – MSC (BHE)	246	320
Average capacity of the BSC – MSC (Mbit/s)	34	34
Average length of the MSC – MSC (Kms)	150	150
Average capacity of the MSC – MSC (BHE)	1,476	1,848
Average capacity of the MSC – MSC (Mbit/s)	155	155 + 34
BHCA per MSC (000)	300	300
Annual minutes served by each MSC (mln)	1,500	1,500
BHE per MSC port	15	15

* *The numbers assumed are illustrative. The source for some of the numbers is the Oftel 1998 model.*

A1.2.3 Deriving annual call minutes

To estimate the number of call minutes that an element serves in a year, we have made assumptions about the number of BHE the same element serves. It does not matter whether BHEs act as a capacity constraint; the exercise is not concerned with whether the element's cost driver is call minutes.

The number of BHEs an element serves is then converted into the total number of call minutes carried through appropriate assumptions. The final results are going to be quite sensitive to the assumptions made about the relationship between annual call minutes and BHEs. The conversion factor is here defined through the following equation:

$$\text{Conversion factor} = \text{Annual call minutes} / \text{BHE}$$

Converting BHEs into annual call minutes is the reverse exercise to that usually undertaken in bottom-up cost models, where annual traffic (along with reasonable assumptions on margins for growth) is converted into BHE for the purposes of dimensioning the network. When converting annual minutes in BHE (or call attempts in BHCA), the modeller usually takes into account deviations of weekly traffic over the year, of daily traffic over the week and of busy hour traffic over the day. This is done to smooth the effect of unreasonably busy weeks of the year, days of the week and hours of the day. Average deviations are taken into account and the traffic over the Busiest Hour of the year (BHE) worked out accordingly.

Theoretically, when doing the reverse operation, the same factors should be taken into account. To do this correctly would require country-specific information.

For the purposes of this exercise we refer to the conversion factor used in the publicly available Analysys model developed for Oftel in 2001. The used conversion factor and the assumptions underlying it are summarised in the following table.

Table A2: Assumptions underlying BHE factor

	Analysys assumption
Percentage of daily calls attempted during the Busy Hour	15%
Number of Busy Hour days in a year	250
Average duration of a call attempt (minutes)	3
Conversion factor (i.e., No of annual call minutes per BHE)	100,000

Source: Analysys model developed for Oftel in 1998

This factor is lower than the one used in the Adaptable Bottom-Up model Europe Economics developed for the European Commission to estimate interconnection costs in the fixed network. We would expect the factor to be lower for mobile networks if a lower percentage of annual calls are made during the Busy Hour than is the case for fixed networks.

The assumed factor has a big effect on results. If the percentage of daily calls attempted during the busy hour was just 10 per cent, then the conversion factor would be 50 per cent higher. As a consequence, the per minute costs estimated would be 33 per cent lower than those reported in our illustrative example here.

A1.2.4 Cost assumptions

The following table summarises the cost assumptions underlying the per minute estimates of the network elements. The source of this information is the Analysys model built for Oftel in 1998.[59]

[59] Figures up to 5 digits have been rounded to the next hundred, figures with 6 digits and more have been rounded to the next thousend.

Table A3: Cost assumptions underlying the estimates of the network elements

Equipment	Capital investment cost (€)[60]	Annualised cost (€)	Associated operating costs (€)	Total annual costs (€)
Site				
Site acquisition and preparation	40,000	6,400	-	6,400
Site lease	-	-	4,800	4,800
3 sectors equip.	182,000	73,000	26,000	98,000
1 sector equip.	125,000	50,000	17,000	67,000
TRX	19,200	6,700	2,700	9,400
Backhaul (2 Mbit/s)	31,200	8,400	4,400	12,800
BSC	800,000	264,000	112,000	376,000
BSC – MSC (34 Mbit/s)	127,800	34,500	17,900	52,400
MSC				
Processor	2,179,000	545,000	305,000	850,000
Software	800,000	200,000	112,000	312,000
Interconnection interface	40,000	10,000	5,600	15,600
Switching support plant	160,000	40,000	22,000	62,000
Building	240,000	36,000	34,000	70,000
Site lease	-	-	24,000	24,000
MSC ports	4,600	1,100	600	1,800
MSC – MSC (155 Mbit/s)[61]	N/A	N/A	N/A	200,000
HLR	640,000	160,000	90,000	250,000
Network management	24,000,000	6,000,000	3,360,000	9,360,000

Source: Analysys model for Oftel (1998)

The underlying cost of capital has been assumed to be 14.25 per cent.

60 Exchange rate used: 1£ sterling = 1.6 Euros (as of 25 September 2001).
61 Europe Economics estimate

A1.2.5 Results

The following two tables summarise the results of this exercise.

Table A4: Per minute costs of the network elements –Rural Network

Equipment	Total annual cost (€)[62]	Annual minutes served[63]	Cost per minute (€/100)
1 sector Cell (Total)	78,600	400,000	19.65
TRX	9,400	310,000	3.06
Backhaul	12,800	400,000	3.20
BSC	376,000	24,620,000	1.53
BSC – MSC	52,400	24,620,000	0.21
MSC (Total)	1,333,000	1,500,000,000	0.09
MSC ports	1,800	1,500,000	0.12
MSC – MSC	200,000	147,600,000	0.14
HLR	249,600	1,500,000,000	0.00
Network management	9,360,000	1,500,000,000	0.01
Total			**28.0**

62 See note 2.
63 See note 2.

Table A5: Per minute costs of the network elements – Urban Network

Equipment	**Total annual cost (€)**	**Annual minutes served**	**Cost per minute (€/100)**
3 sectors Cell (Total)	110,000	3,000,000	3.66
TRX	9,400	400,000	2.35
Backhaul	12,800	3,000,000	0.43
BSC	376,000	32,000,000	1.18
BSC – MSC	52,400	32,000,000	0.16
MSC (Total)	1,333,000	1,500,000,000	0.09
MSC ports	1,800	1,500,000	0.12
MSC – MSC	252,400	184,800,000	0.14
HLR	250,000	1,500,000,000	0.00
Network management	9,360,000	1,500,000,000	0.01
Total			**8.1**

Summing the per minute cost charges produces an average per minute cost charge. This would only accord with the per minute cost charge for a particular service if it is assumed that all services make the same use of the different elements. In practice, different types of calls will make more or less use of the various elements so such summing will not be appropriate to get an average cost per minute for an individual service. For example, on average call termination may involve one more MSc and one more MSC-MSC transmission route than call origination

When valuing the results of this exercise, it is important to bear in mind that the network elements, whose costs have been here estimated, do not cover all network costs. All those network components that are in the network to provide functionality to these elements, such as racks for MSCs or power generators placed in the sites, are not included in this estimate. Our ball-park estimate of the total amount of these costs would add up to 15 per cent of the annualised investment costs of these network elements. Non-network costs have ***not*** been included in this analysis. It is also important to remember that the estimates use data that are three years out of date.

APPENDIX 2: OPERATIONS PROVIDED BY A GSM NETWORK

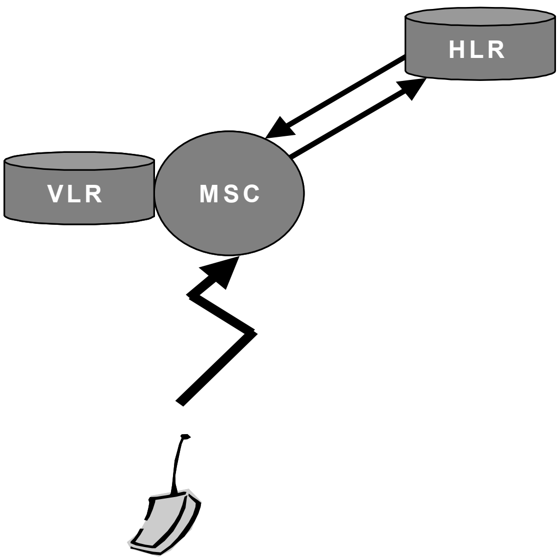
This annex describes the operations provided by a GSM network, outlining the network elements involved in each. The operations categorised are not intended to represent the actual services that subscribers may purchase. Rather, the focus is on the various operations a GSM network undertakes to supply mobile telephony to its subscribers.

A2.1 Location Updating and Authentication

A2.1.1 Location Updates

Location updating is the process by which a user's coverage is maintained as the subscriber moves between different cells and location areas. Since a subscriber must be registered in his/her current location area before telephony services can be enjoyed, location updating is a pre-requisite for making use of the mobile network.

Diagram A1.1: Location Updating



A mobile station that has just been turned on, or ascertains that it has been moved to a different location area,[64] needs to register its new position within the network. This is achieved through the MS sending a signal to the VLR associated with the current location area. The VLR records this information and sends it to the HLR to ascertain whether the user has a subscription with the network. If this is the case, the HLR records the address of the VLR at which the MS is registered and returns administrative information about the subscriber to the VLR (cancelling such information on the VLR the subscriber was previously registered with). The network can then process the subscribers' outgoing calls without the necessity of interrogating the HLR each time because subscriber information has been temporarily copied from the HLR to the visited VLR.

[64] A mobile handset that is switched on monitors the Broadcast Control Channel. The signals it receives through this channel will depend on the location area the handset is in, and hence the handset is aware when it enters a different location area.

(For reliability purposes, each switched-on MS will usually update its location periodically regardless of whether or not there has been change in location.)

A2.1.2 Authentication

In addition to the exchange of subscriber information between HLR and VLR, the location updating procedure is also used for network security purposes. By enquiries with the AuC and EIR, the network can ensure that services are restricted to legitimate subscribers and that no blacklisted (e.g. reported stolen) handsets access the network. This authentication takes place whenever an MS tries to connect to the network, so occurs for location updating as well as call origination and termination.

A2.2 Voice Calls

The primary use of a GSM network is to convey voice traffic to and from subscribers. Call origination refers to the connection established between the calling party and the point of interconnection between the calling party's network and the receiving party's network. Call termination refers to the connection established between a point of interconnection and the receiving party. A GSM network operator controls both origination and termination of calls across its own network.

Below we trace through the basic procedure for originating and terminating a call on a mobile network. It is instructive to split a mobile telephony call into three (potential) operations: call set-up, voice conveyance and handover. Call set-up is the initial attempt to establish a connection between the two parties involved in the call. If this is successful a voice channel is established between the parties; this is the voice conveyance operation. Thirdly, handover is required if a mobile subscriber who is making or receiving a call moves between cells.[65]

A2.2.1 Call set-up

The process for establishing a connection between caller and receiver requires use of signalling. The signalling procedure across the mobile network for call origination differs to the procedure when the network is terminating a call. The two are outlined separately below and then contrasted to illustrate the key differences.

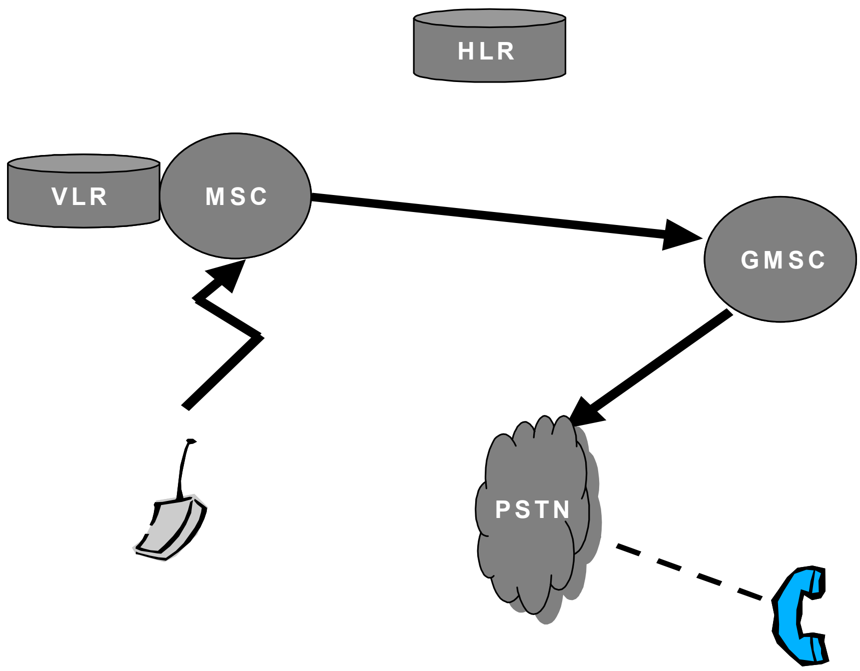
A2.2.1.1 Call Origination

When a subscriber makes an outgoing call, the handset sends a signal (containing information on the user and the call attempt) to the BTS that appears to offer the best coverage in the area; this is not always the nearest BTS. The signal is passed via the BSC to the MSC/VLR associated with the location area the subscriber is in. If the subscriber is currently registered with that VLR (i.e. the user's current location has been updated) the information and permission necessary for the call to be made on the network is contained on the VLR and the signal proceeds through the MSC layer to an appropriate GMSC (for calls that are to another network, mobile or fixed). For

[65] A further operation, billing is described separately as a service in its own right. It need not apply solely to voice calls.

calls identified as on the same network, the MSC then effectively plays the role of the GMSC in call termination.

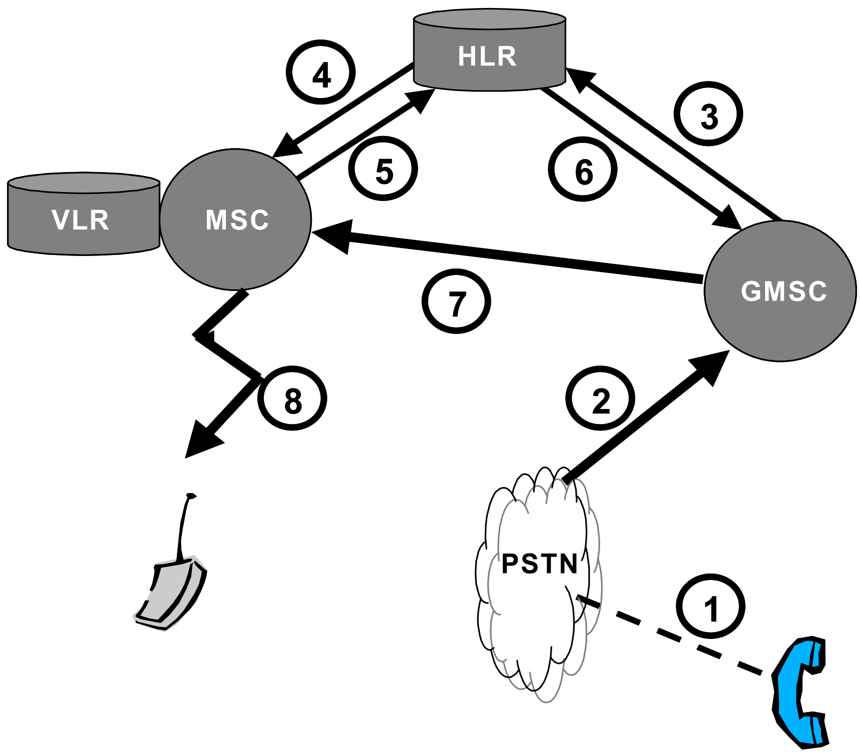
Diagram A1.2: Call Origination



A2.2.1.2 Call Termination

In order for a call to be terminated on a mobile network, the subscriber to be contacted is temporarily allocated a Mobile Station Roaming Number (MSRN). The call is routed by matching the (temporary) MSRN to the subscriber's (permanent) IMSI. The steps involved are described below.

Diagram A1.3: Call Termination



1 A caller from another network dials a mobile number

2 For a call from the fixed network, the call is routed through the PSTN and a signal sent to a GMSC of the terminating network.

3 The GMSC interrogates the HLR for the appropriate MSRN.

4 The HLR does not usually hold the MSRN but does hold the address of the VLR the subscriber is currently registered with. Hence the HLR enquires with this VLR.

5 The VLR allocates a temporary MSRN for the subscriber to be contacted.

6 The HLR sends this back to the GMSC.

7 The GMSC then contacts the VLR

8 The mobile station with the IMSI corresponding to the MSRN is informed of the incoming call.

Call set-up can be thought of as two operations when dealing with call termination. The first is to locate the subscriber while the second is to inform the subscriber of the incoming call. The former stage is an information exchange between the GMSC, the HLR and the subscriber's current VLR (stages 3–6 in Diagram A1.3). Notifying the subscriber of the call is essentially transmission of information from the GMSC through the appropriate MSC and BSC to the BTSs in the subscriber's current location area.

A2.2.1.3 Difference in setting up originating and terminating calls

Diagrams A1.1 to A1.3 show how the network is used for location updating, and the setting up of call origination and call termination. The diagrams highlight the information exchanges that take place, rather than the full routes that signals travel. Hence, although any communication between a VLR and a HLR will necessarily involve at least one MSC, Diagram A1.3 is a simplification because the transit from VLR to HLR is ignored.

In terms of call set-up, termination differs from origination because of the initial stage of obtaining information from the HLR as to which MSC/VLR the subscriber is currently associated (stages 3–6 in Diagram A1.3). The common procedure in call set-up for origination and termination is the communication between the MS (via BTS, BSC, and MSC) and the GMSC, which requires information from the VLR. In the case of origination this information pertains to the services the subscriber has access to, while for termination the information concerns the current location area of the subscriber. In both cases this information has been transferred to the VLR in location updating.

A2.2.2 Voice Traffic

Regardless of whether a mobile network is originating or terminating a call — once it has successfully been set-up — a voice channel will be established between the GMSC and the subscriber's handset, via BTS, BSC and MSC(s). Although such a GMSC to MS connection is required for both origination and termination, the particular GMSC used and the route taken across the mobile network may differ between the two.

Where there are multiple PoIs between the originating and terminating networks, efficient routing of the call would require the originating network to have knowledge of which PoI is most appropriate to send the call through. However, only the terminating network has access to its HLR and VLRs. Consequently information regarding the location of the mobile handset is unavailable to the originating network. Since the originating network cannot ascertain the most appropriate PoI for the call, the originating network should route the call to the PoI that minimises its own transit costs.

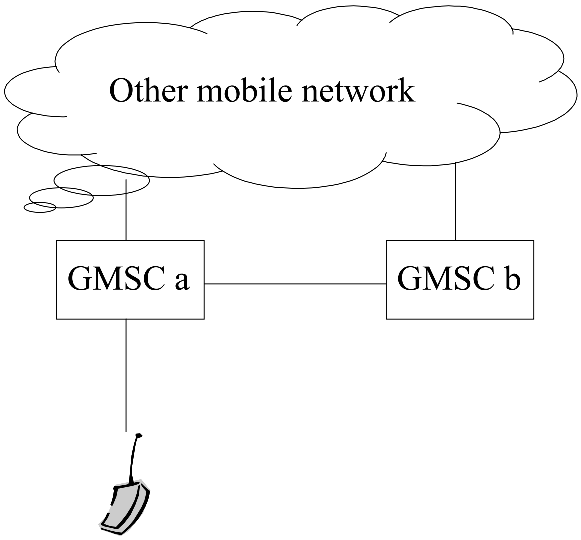
By contrast, when a call is originated on a mobile network, the network provider is able to choose the PoI through which the call leaves the network. Because the cost of routing the call from the handset to a point of interconnection will vary between different interconnection points, the mobile operator can reduce costs by selecting the PoI leading to the least-cost origination.[66]

[66] If the call is to be terminated on a fixed network, it is possible that the mobile operator can take account of which POI will result in the lowest termination charges because the mobile operator can ascertain the (fixed) end point of the call. If the call is to be terminated on another mobile network, a flat fee is charged for terminating a call anywhere on the network, so the termination charge is not affected by the PoI used.

A2.2.2.1 Mobile-Mobile Calls (Off-Net

Diagram A1.4 below highlights how, with multiple interconnection points to a mobile network, the signalling and voice traffic are likely to take a longer route (on average) through the network for call termination than origination.

Diagram A1.4: Off-Net Mobile-Mobile Call)



When a call is originated from the MS, it travels from MS to MSC via the BTS and BSC. The network operator can choose to route the call through the interconnection point at GMSC a or the interconnection point at GMSC b. Since the operator will face a flat fee for termination the origination cost will be minimised by routing the call through GMSC a.

When a call is to be terminated on this mobile network, the network operator has no choice over whether the call is routed through GMSC a or GMSC b. Suppose that half the time the call comes through GMSC b, which is a less direct route. The higher costs of using this less direct route imply that on average termination costs be higher than origination costs (where the more direct route is always used)[67].

The more interconnection points there are between the two networks, the greater is the potential to reduce origination costs because the average distance from each BTS to POI falls. Therefore, the more interconnection points there are, the higher will be the (average) cost differential between call termination and call origination.

A2.2.2.2 Mobile-Fixed and Fixed-Mobile Calls

As for the mobile-mobile case, when a mobile network terminates a call from a fixed network it does not have choice over which GMSC receives the call. In Diagram A1.5, because the fixed

---

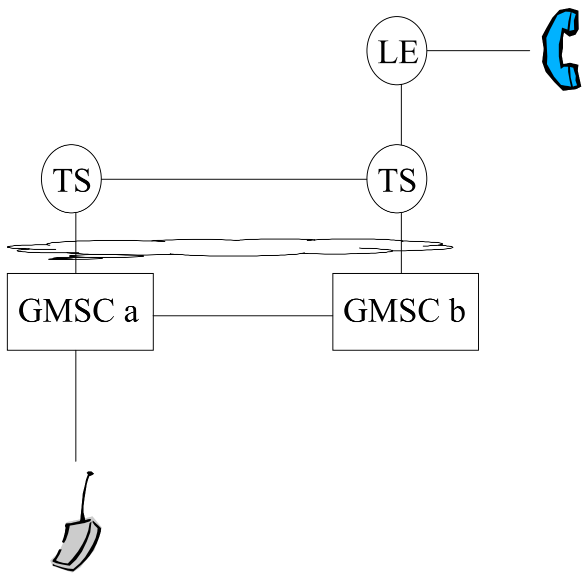
67 Note that if the more direct route is very congested, it may be less costly to route the call through GMSC b. The important point is that the mobile operator has more control over the routing and therefore costs of call origination.

operator does not know where in the mobile network the MS is located, it will send the call to GMSC b, the PoI closest to the fixed phone originating the call.

For a mobile-to-fixed call, the charge for terminating the call on the fixed network does tend to vary with the location of the party being called (which the mobile network operator knows because the PSTN has a geographical numbering scheme). In order to minimise the total costs of the call, the mobile network operator will choose the PoI that minimises the sum of the costs of call origination (across its network) and the charge for call termination (across the fixed network). Thus in Diagram A1.5 the mobile network operator may choose not to route the call through the GMSC nearest to the caller. Instead the MNO may choose to originate the call via GMSC b, and thus only have to pay for single transit in the fixed network rather than double transit. However, normally MNOs will originate calls out of their network via the most direct route even for calls to fixed networks since the costs of transit will typically be cheaper in other networks.

As for mobile-to-mobile calls, the more interconnection points there are, the higher will be the (average) cost differential between call termination and call origination. This is because the MS will have shorter routes to the nearest PoI, while the probability of an incoming call entering the network via the same PoI is reduced.

Diagram A1.5: Fixed-Mobile Calls



A2.2.3 Handover

As a subscriber in dedicated mode moves from cell to cell the network needs to ensure a smooth handover. A handover can be initiated by the handset or the MSC as a means of traffic load balancing. At least once per second the handset will pass information to the BSC and MSC concerning the best candidates for handover, based on received signal strength. There is not a recommended algorithm for GSM concerning when a handover should take place.

Internal handovers involve transferring the call to a different channel (time slot) in the same cell or to a different BTS but under the control of the same BSC. Such handovers are handled by the

BTS, and the only signalling that goes to the MSC is to notify it of completion of the handover. *External handovers* are handled by an MSC. They involve the transfer of calls between cells that are controlled by different BSCs, whether or not the same MSC controls the BSC. The original MSC (or *anchor MSC)* remains responsible for most call-related functions, with the exception of subsequent external handovers that the *relay MSC* will handle.

In effecting a handover there are no differences between call origination and call termination.

A2.3 Billing

Billing can be thought of as two distinct stages:

1 Recording a subscriber's usage of network services

2 Using this information to calculate charges and invoice the subscriber

The first stage is undertaken at the anchor MSC that handles the subscriber's voice call or other network usage. The MSC creates a Call Detail Record (CDR) after each call (or during a call if it is of a long duration) which records information such as the user's IMSI and current location area as well as the source, destination and duration of the call.

The second stage of billing converts this record of network usage into a charge against the subscriber's account (or, for call termination, the account of the originating network, domestic or foreign). For charges that accrue to the network's own subscribers, how this is undertaken will depend on whether the subscriber is a prepaid or post-paid customer.

For a post-paid customer, the CDRs are transferred from the MSC to a central billing system for processing. Charges can be calculated by applying the subscriber's tariff structure to the usage recorded in, say, the latest month's CDRs pertaining to the subscriber, collated from all MSCs in the network. The subscriber can then be invoiced for the month's call charges

By contrast, a prepaid subscriber purchases a certain amount of call credit before making use of the network's services. Then, as the subscriber originates calls and builds up charges, these are set against his/her call credit. This can be done through temporary accounts in the HLR using IN with real-time (or hot) billing.

For roaming (post-paid) the CDRs generated by visited network MSCs are usually forwarded in bulk to home network for payment. Under roaming rules, TAP files are sent at least once very 24 hours.

A2.4 Other Operations

Mobile networks are capable of a number of other procedures. For example, call waiting notifies a subscriber of an incoming call during a conversation.

A2.4.1.1 Short message service

SMS allows subscribers to send and receive text-based messages of up to 160 characters in length. These messages are sent using narrow bandwidth. Delivery speed can vary according to the volume of traffic on the network. The service is provided via a Short Message Service Centre (SMSC) that operates a store-and-forward system in which each subscriber has his/her own mailbox. Once received at the network's SMSC, a SMS message is forwarded to the recipient's handset. If the recipient cannot be reached (e.g. the handset is switched off) the message is stored at the SMSC, which will attempt to deliver the message later.

International messages are feasible because SMS messages can be sent to any MS on a GSM network.

A2.4.1.2 Voicemail

When a call is to be terminated on a mobile network, a signal is sent to the handset in order to alert the receiving party to the incoming call. If the handset is switched off, or if the user does not answer this call, the calling party may be re-directed to the network's voicemail system, where a (voice) message can be left. The voice channel that is set up over the mobile network is therefore one from the GMSC to the network's voicemail system.

A2.4.1.3 Data services

There are a number of data services that MNOs offer on their GSM networks. The Wireless Application Protocol (WAP) has become a de-facto standard for the presentation of information on mobile telephones. WAP content is hosted on WAP servers or gateways. To connect to the Internet, a WAP-enabled handset will transfer data through the MNO's network, over the PSTN and then to and from the Internet through a remote access server. Once on the Internet the connection must go via a WAP gateway so that the protocols can be changed.

A number of 2G+ or 2.5G technologies have been developed that seek to offer higher data rates, such as high-speed circuit switched data (HSCSD) and general packet radio services (GPRS). HSCSD works on the current GSM physical network; it offers higher data rates by using more than one time slot for single connection. Since HSCSD is a circuit-switched connection, it is similar in its network usage to voice traffic, the difference being that a minute of HSCSD traffic uses more channels than a minute of a single voice call.

In contrast, GPRS is a packet-switched technology. It requires new service nodes and new control channels to form the packet switched core network that GPRS uses. The initial connection still uses the BTS and BSC of the GSM network, but from the BSC the packets are transferred to the GPRS-support nodes instead of the MSC where a voice (or HSCSD) call would go.

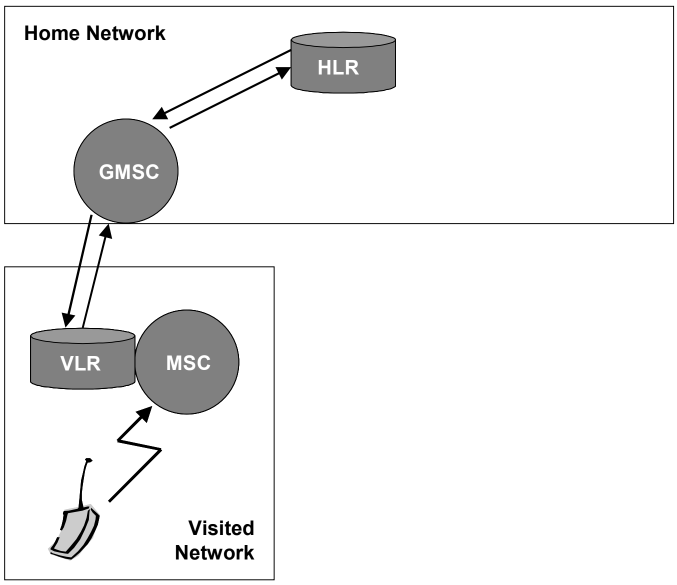
Enhanced Data rates for GSM Evolution (EDGE) will allow even higher data rates, but this technology will require new handsets and a slight re-engineering of the GSM network since a single BTS could not transmit GSM and EDGE carriers.

A2.5 Roaming

A2.5.1 Location updates when roaming

When a subscriber to one mobile network is roaming on another mobile network (the visited network) the subscriber's MS must update its location with the visited network. This is undertaken through the MS sending a signal to the visited VLR serving the location area the roaming subscriber is currently in. The visited VLR identifies the home country and network that the user has a subscription with from the user's IMSI, and hence knows which network's HLR needs to be queried in order to obtain information on the subscriber. If the subscriber has roaming rights and permission to access the visited network, subscriber information will be copied from the home network HLR to the visited network VLR. This enables the roaming user to send and receive calls via the visited network. Additionally, the home network HLR will note the address of the VLR that the subscriber is registered with. When faced with an incoming call, the HLR is aware that the subscriber is roaming on another operator's network and can contact the appropriate VLR on this network.

Diagram A1.6: Location Updating for Roaming Subscriber



A2.5.2 Making a call when roaming

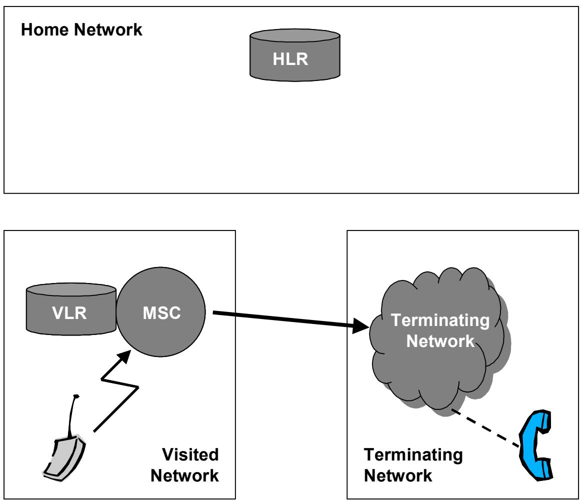
The information necessary to enable the roaming subscriber to make and receive calls has been copied from his/her home network HLR to the visited network VLR in location updating. Therefore, when the roaming subscriber makes a call, the call origination procedure is identical to the case where the visited network operator is originating the call from one of its own mobile subscribers. The costs of originating the call should therefore be the same as if a corresponding call was made by one of the visited network's own subscribers. The corresponding calls that we might compare with different types of roamed calls are identified below.

- *In-country calls.* There is no difference in terms of call routing between a network originating an in country call from a roaming subscriber and originating a domestic call from one of its own subscribers. The origination routing is as discussed before.

- *Calls home.* A visited mobile operator originating a call home by a subscriber roaming on its network will undertake the same procedure as if it was originating an international call from one of its own subscribers to a fixed or mobile network in the roaming subscriber's home country.

- *Country-to-country calls.* Originating a country-to-country call from a roaming subscriber is identical to the visited network originating an international call (to that country) from one of its own subscribers.

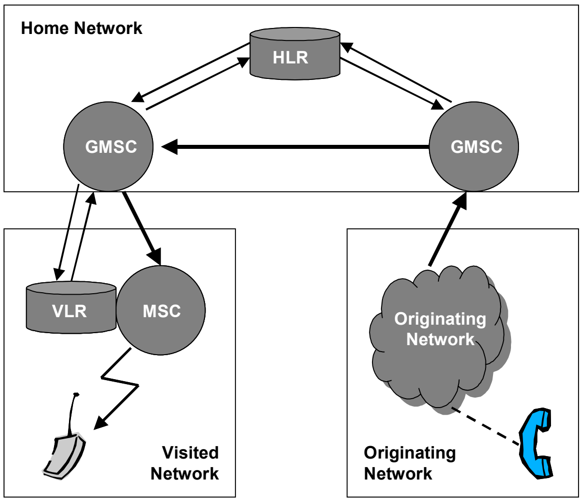
Diagram A1.7: Roaming subscriber originating a call



A2.5.3 Receiving a call when roaming

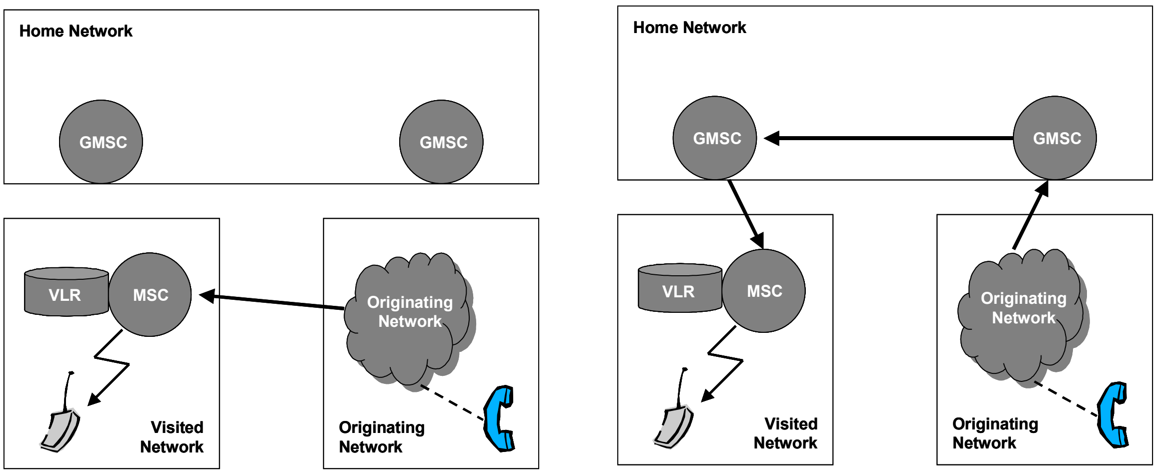
A call to a roaming subscriber will be initially directed to the subscriber's HLR on his/her home network. The HLR is aware (from location updating) that the subscriber is currently roaming and obtains an appropriate MSRN from the visited network. The MSRN is returned to the home network GMSC and the call proceeds to the visited network. The visited network notifies the roaming subscriber of the incoming call. If the call is accepted, a connected is established from the originating caller's network to a PoI with the subscriber's home network and from there, via a PoI with the visited network, to the roaming subscriber. This is shown in the following diagram.

Diagram A1.8: Terminating a Call to a Roaming Customer



In the case of international roaming, there exists potential for indirect routing of a call to a roaming subscriber. The diagrams below ignore the signalling involved in call set-up and concentrate on the route the call takes between the two calling parties.

Diagram A1.9: Alternative Call Transit Routes for Call Termination



Bypassing the home network *Call transit via the home network*

There are three networks involved in the call (ignoring any network that undertakes transit on behalf of one of these). The home (mobile) network of the roaming subscriber, the visited (mobile) network that the subscriber is roaming on and the originating (mobile or fixed) network that the originating caller is connected to. PoI A provides interconnection between the originating network and the subscriber's home network. PoI B provides interconnection between the subscriber's home network and the visited network.

The table below summarises, for these three call types, the networks that would be in the same country and those that would be in different countries.

	Networks in the same country	Networks in different countries
In-country call	originating & visited networks	home network
Home country call	originating & home networks	visiting networks
Country-to-country call	-	originating, home & visited networks

- *In-country calls.* Although there is no international dimension to the conversation that takes place, an in-country call is international in terms of call routing. The originating and visited networks are in the same country, but the call is international in nature because the call is routed between these two networks via the home network (which is in a different country). Both PoI A and PoI B signify international gateways, hence the call will use two international legs to carry the voice traffic.

- *Calls home.* When a call to a roaming subscriber is originated in the subscriber's home country the call only has one international leg, that from the subscriber's home network to the foreign network the subscriber is roaming on. This is because POI A is an interconnection point between networks in the same country.

- *Country-to-country calls.* A country-to-country call will have two international legs; both POI A and POI B signify international interconnections. All three networks involved in the call are in different countries.

Only the originating and visited network are essential for the voice transit (i.e. excluding set-up and signalling) of the call. Once the home network HLR has located the visited network VLR and retrieved the MSRN, the call is just conveyed across the home network from one GMSC to another. It is potentially more efficient if the call travels directly from the originating network to the visited network, bypassing the home network altogether.[68]

This bypass is likely to be of greatest advantage for in-country calls because the voice transit will then be entirely domestic. Even for a country-to-country call the saving due to more efficient routing could be large. For example, consider a call from France to an American GSM subscriber roaming in Germany. If the voice traffic is conveyed through the subscriber's home network in the USA, two transatlantic legs are required. Whereas if the traffic travels directly from the French to German network, costs for conveying the call could be significantly lower.

The costs associated with transit of the call will therefore depend on the technologies and agreements that are in place. If the home network can be bypassed, clearly the costs associated with a call to a roaming mobile customer will be lower than if such bypass is not possible. Even if it is technically feasible, an MNO may not use the technology, possibly because a rival has

68 Note that for calls that originate from a roaming subscriber, the subscriber's home network is not involved in the call, which passes direct from the visited network to the receiving party's network.

proprietary rights to it. In a regulatory context, forward-looking cost models will need to make an assumption about whether bypass is possible. Once CAMEL Phase II is adopted such bypass will be feasible.

APPENDIX 3: ABBREVIATIONS USED

AMPS	Advanced Mobile Phone Service
AuC	Authentication Centre
BHE	Busy Hour Erlangs
BSC	Base Station Controller
BTS	Base Transceiver Station
CAMEL	Customised Applications for Mobile networks Enhanced Logic
CRD	Calling Record Data
EDGE	Enhanced Data rates for GSM Evolution
EIR	Equipment Identity Register
ETACS	Extended Total Access Communication System
FAC	Fully Allocated Costs
FDC	Fully Distributed Costs
FL-LRIC	Forward Looking Long Run Incremental Costs
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GSM[69]	Global System for Mobile communication
HLR	Home Location Register
HSCSD	High-Speed Circuit-Switched Data
IMSI	International Mobile Station Identity
IN	Intelligence Network
IOT	Inter-Operator Tariff
ISDN	Integrated Services Digital Network
LRIC	Long-Run Incremental Cost
Mbit/s	Mega bits per second
MNO	Mobile Network Operator
MS	Mobile Station (the handset)
MSC	Mobile Switching Centre
MSISDN	Mobile Station international ISDN
MVNO	Mobile Virtual Network Operator
NNT	Normal Network Tariff
NPV	Net Present Value
NRA	National Regulatory Authority
NRV	Net Realisable Value
PoI	Point of Interconnection
PSTN	Public Switched Telephone Network
RC	Replacement Cost
SMP	Significant Market Power
SMS	Short Message Service
SMSC	Short Message Service Centre
STIRA	Standard International Roaming Agreement
TAP	Transfer Application Protocol

69 In this report, GSM has sometimes been used when referring to second-generation mobile networks.

TSC	Transit Switching Centre
VLR	Visitor Location Register
VPLMN	Visited Public Land Mobile Network
WACC	Weighted Average Cost of Capital
WAP	Wireless Application Protocol

APPENDIX 4: ACKNOWLEDGEMENTS

During this study we have benefited from meetings, e-mail correspondence and telephone calls with a number of interest parties. We thank the following for their help, without implicating them for the conclusions reached.

Agcom
ART
Blu
BT Concert
BT Wireless
Cable and Wireless
Comfone
EETT
FICORA
INTUG
MCI Worldcom
Mint Telecom
NPT
Oftel
Omnitel
One2One
OPTA
Orange
TIM
PTS
RegTP
Sense
Telestyrelsen
Versatel
Vodafone
Wind

APPENDIX 5: SOURCES FOR PRICE INFORMATION

Current price data is available from the mobile companies websites. These do not provide time series data. The price of baskets can be calculated from them, an exercise that various studies have undertaken. For their own markets most regulators have undertaken some work collecting and analysing data. Not all of this is in the public domain. It is also difficult to use such data to make cross-country comparison.

Many reports relating to the mobile industry quote the same price data. Some of the more common sources for such data are listed below.

Teligen *(www.teligen.com)*

Teligen was set up in 1999 to carry out the commercial activities of Eurodata Foundation In June 1999 it produced the "Report on Telecoms Tariff Data" for EC Directorate General XIII.

Through a special agreement with the OECD, Teligen offers a commercial telecommunication tariff comparison product. Entitled T-Basket, the product enables users to compare tariffs updated on a quarterly basis. Results are available for the following services: national public switched telecommunication networks (business and residential PSTN); international PSTN; mobile telephony; X25 packet switching; and national leased lines.

EMC *(www.emc-database.com)*

EMC use mainly primary sources of information in their research methodology, such as correspondence and contacts within companies worldwide.

The EMC database reportedly contains cellular tariff information; comparative views in common currency ($US); prepaid tariffs; call charges; monthly rates; connection charges; bundled minutes of use; value-added services and costs; text notes for each country. The information can be viewed by country, by region, by technology, by date and usage scenario.

Fisher & Lorenz *(www.fl.dk)*

Fisher & Lorenz have collected data on behalf of INTUG, who have published two reports on international roaming charges since 1999.

Ovum *(www.ovum.com)*

The OECD quote quarterly figures from Ovum concerning mobile termination rates.

Tariffcentre (www.tariffcentre.com)

Have retail prices for fixed-to-mobile calls for all incumbent fixed operators in all EU countries barring Finland and Greece.