# Mobile Call Termination

Armstrong, Mark and Wright, Julian

Department of Economics, University College London

July 2008

# Mobile Call Termination*

Mark Armstrong† and Julian Wright‡

July 2008

**Abstract**

Motivated by recent UK experience, we study the problem of mobile call termination. This is an intriguing policy story, in which regulation has been imposed on what appears to be a competitive industry. We introduce a framework which integrates two existing literatures: one analyzing calls from the fixed network to mobile networks (where the predicted market failure involves the termination charge being set at the monopoly level), and one analyzing calls from one mobile network to another (where the predicted unregulated termination charge lies below the efficient level). Our unified framework allows us to consider the impact of wholesale arbitrage and demand-side substitution. In the absence of very significant market expansion possibilities, we find the unregulated termination charge lies *between* the efficient and the monopoly benchmarks. There remains a rationale for regulation, albeit reduced relative to the earlier literature.

# 1 Introduction

There is an important set of markets in which monopoly prices emerge even when competition is intense. That is to say, while industry profit is not excessive overall, there is an inefficient balance of prices: too high for some services, too low for others. Familiar examples involve consumer "lock-in" of various kinds, including markets with switching costs.[1] In these markets the typical pattern of prices involves "bargains-then-ripoffs", so that firms attract new consumers with generous deals up-front and consumers pay high (perhaps monopoly) prices once locked in. If competition is vigorous, the monopoly profits from locked-in consumers are transferred to new consumers, and the lifetime profitability of a consumer is approximately zero. In a sense, a consumer's "future self" is exploited by her "present self".

*We are grateful to Stefan Behringer, David Harbord, Bruce Lyons, Paul Muysert, David Sappington, Tommaso Valletti, John Vickers, Ingo Vogelsang, Helen Weeds and to two referees for very helpful comments. Armstrong is grateful for the support of the Economic and Social Research Council (UK) and Wright gratefully acknowledges the support of the Singapore Ministry of Education AcRF Tier 1 fund under Grant No. R122000080-101/112/133.

†University College London, email mark.armstrong@ucl.ac.uk.

‡National University of Singapore, email jwright@nus.edu.sg.

[1] See Farrell and Klemperer (2007) for an overview.

A related, but distinct, set of markets exhibit what might be termed "competitive bottlenecks". Here, firms compete to attract one group of consumers. For natural technological or geographical reasons, each consumer in this group wishes to deal with just one firm. A second group of consumers wishes to interact with the first group and, because each consumer in the first group deals exclusively with one firm, that firm can charge the second group high prices (or pay low input prices) for access to its captive customers. If competition is vigorous, the monopoly profits generated by the first group are passed back to these consumers in the form of subsidised service. Here, the first group may be said to exploit the second. Examples of this kind of market include: newspaper advertising (where most readers tend to read a single newspaper due to time constraints, and so a newspaper can charge high fees to advertisers for access to its captive readers) and supermarkets (where a consumer tends to visit just one shop over the relevant time horizon, and so suppliers are prepared to be paid low input prices for access to these captive shoppers). At the end of the paper we will discuss another possible application of this approach, to the contentious issue of "net neutrality" on the internet.[2]

However, perhaps the leading example of a competitive bottleneck is call termination on mobile telephone networks. Call termination refers to the service whereby a network completes—or "terminates"—a call made to one of its subscribers by a caller on another network. We focus on mobile, as opposed to fixed-line, networks since the former markets are currently much more competitive than the latter in most countries, and the issue of regulation there is more controversial. Mobile networks compete for subscribers, but in the absence of regulation they may charge other networks excessively to talk to their subscribers. Concerns about mobile call termination being a bottleneck, and the associated high charges for calling mobile subscribers, have led to regulation of termination charges around the world. For example, caps on mobile call termination charges are applied throughout the European Union. In some countries, notably the United States, mobile termination charges are indirectly regulated (at low levels) through reciprocity requirements with the fixed-line networks.

There are two broad types of call termination on mobile telephone networks: termination of calls made by callers on the fixed-line telephone network (fixed-to-mobile, or FTM, termination) and termination of calls made from other mobile networks (termed mobile-to-mobile, or MTM, termination).[3] In the literature to date, FTM and MTM termination have been largely treated separately, with distinct market failures highlighted. One aim of this paper is to integrate these two strands.

Broadly speaking, FTM call termination viewed in isolation is likely to involve unilateral monopoly pricing if left unchecked. The vast majority of mobile subscribers join just one mobile network, and so callers on the fixed telephone network must route calls through a subscriber's chosen network. No matter how competitive the market for mobile subscribers may be, a mobile network holds a monopoly over—and can set high charges for—delivering calls to its subscribers. That is to say, as was shown formally by Armstrong (2002, section 3.1) and Wright (2002), FTM call termination gives rise to a competitive bottleneck and

[2] See Armstrong (2006) and Armstrong and Wright (2007) for general analyses of competitive bottlenecks.

[3] Text messaging (or SMS) termination fits into our framework of MTM call termination although in most countries it is not yet regulated.

regulators need to worry about excessive termination charges.

The study of MTM termination (often termed "two-way interconnection" in the literature) has led to quite a different focus: whether mobile networks can use a negotiated termination charge to relax subsequent competition for subscribers.[4] They can do this by setting MTM termination charges below cost. This causes off-net calls (i.e., calls from one mobile network to another mobile network) to be cheaper than on-net calls (calls from one mobile network to the same network), so that consumers prefer to join the smaller network. This in turn reduces each network's incentive to attract subscribers. Thus, when networks can coordinate on the choice of a MTM termination charge, they may have an incentive to choose too *low* a charge. This low termination charge acts—somewhat counter-intuitively—to harm mobile subscribers, who prefer the intense competition which accompanies a higher charge. This analysis gives rise to a puzzle. Even though the theory predicts that networks will want to set MTM termination charges that are too low, to the best of our knowledge no regulator has taken seriously such a concern. Rather, as with FTM termination, policy has acted to prevent firms choosing high MTM termination charges.

In section 3 we present a benchmark model in which both FTM and MTM calls are present but where the two termination charges are determined separately. This integrates the existing literature and captures the opposing incentives networks face in their determination of the two different charges. In section 4 we show this puzzle about MTM termination can be reconciled by taking into account the practical constraint of wholesale arbitrage. By this we mean that a mobile operator cannot maintain a high FTM termination charge together with a low MTM termination charge, since the fixed network could then "transit" its calls via another mobile operator and so end up paying the lower MTM rate (plus a small transit charge). As a result, a mobile operator is forced to set (approximately) a uniform termination charge for FTM and MTM traffic. We establish that this requirement can, by itself, make sense of the regulatory concern that unregulated charges are too high rather than too low.

If the networks collectively set a uniform termination charge, their preferred charge will reflect the relative importance of FTM and MTM termination. If the former is more important, firms will want to set a uniform termination charge which is too high rather than too low (section 4.1). If instead networks choose their uniform termination charges unilaterally (as perhaps is more reasonable), then the temptation to extract termination profits always dominates the incentive to set a low termination charge in order to relax network competition. The result is that unregulated termination charges will be set above the efficient level.[5] Nevertheless, a network's incentive to set monopoly termination charges is mitigated. Setting the uniform termination charge too high strengthens network effects, making the firms tougher rivals. To avoid this effect, networks will keep their termination charges below the monopoly level (unless market expansion possibilities are very significant).

Additional effects arise when we take into account that fixed-line callers are often also mobile subscribers, so that the two groups are not disjoint as assumed in the standard competitive bottleneck story. Such callers can make MTM calls instead of FTM calls if the former

[4] See Laffont, Rey and Tirole (1998b) and Gans and King (2001).

[5] This resolution of the puzzle fits what happened in France and New Zealand, where unregulated networks initially agreed on bill-and-keep (or zero charges) for terminating MTM calls while at the same time setting high FTM termination charges. At some point these very different charges became impossible to sustain, and networks moved to uniform FTM and MTM termination charges but at above-cost levels.

are cheaper. In section 4.3 we find this demand-side substitution between FTM and MTM further weakens the competitive bottleneck of call termination and brings the equilibrium charge closer to the efficient level. As a result, the welfare gains from regulation are smaller than the previous literature may have indicated, although the predicted termination charge without regulation remains excessive. Another important impact of demand-side substitution is to reduce the need to regulate the retail markup on FTM calls, since market power here is constrained by the competing MTM service.

Finally, section 5 briefly discusses the robustness of our results to vertical integration between fixed and mobile networks, with section 6 providing some concluding remarks. Before turning to the analysis, though, we review some relevant aspects of the UK mobile sector and its regulation. This will be used to motivate our particular theoretical framework and to discuss the policy implications of our results throughout the paper.

# 2 UK Experience

Since their inception, regulatory price caps on mobile termination in the UK have faced considerable resistance from mobile operators (as well as support from fixed-line operators). They have also resulted in a large amount of paperwork. Since 1997, there have been several thousand pages of publicly released government reports and submissions on mobile call termination in the UK. Using materials from these reports, in this section we give a brief overview of the most relevant aspects of the UK mobile market.

The regulation of mobile termination dates back to at least 1998 when Oftel (then the UK telecommunications sector regulator) tried formally to control the FTM termination charges set by the two largest mobile networks *Cellnet* (the precursor to the current $O_2$) and *Vodafone*. The proposed regulation was challenged by these mobile operators, leading to a competition enquiry—see MMC (1999). This enquiry did not investigate MTM termination charges, nor did it investigate FTM termination charges levied by the two newer networks, *Orange* and *T-Mobile*, which had entered only recently at that point. The enquiry concluded that the two established networks' FTM termination charges were too high in relation to cost, and based on its recommendations, Oftel regulated their FTM termination charges with a price cap. The imminent expiry of this price cap led to a 2002 Competition Commission enquiry—see Competition Commission (2003). The Commission upheld Oftel's proposed regulation, which covered all four mobile operators and both FTM and MTM (voice) termination charges. (However, SMS termination was not covered in the enquiry, nor by subsequent regulation to date in the UK.) Shortly before the 2002 enquiry, a fifth network (*H3G)* entered the market, although this immature network was excluded from that investigation.

Subsequent reviews by Ofcom (the current UK telecommunications regulator), in 2004 and 2007, extended these regulations. As of 2007, all five networks are subject to regulated price caps for call termination, with reductions in these caps applying from 2007 through to 2011, at which point they will be reviewed again (see Ofcom (2007b)). Table 1 gives an indication of the history of average mobile termination charges in recent years (for all UK networks).

Table 1: Average mobile termination charges (in pence per minute)[6]

| | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|
| all UK operators: | 11.1 | 10.7 | 9.9 | 7.9 | 5.9 |

Thus, termination charges approximately halved over five years, in large part due to tightened regulation. However, the networks do not all set the same average level of charges, as shown in Table 2 which reports the termination charges set by the five operators in 2006. In 2006, the newest entrant, *H3G*, faced softer regulation than the established networks, and took advantage of this to set termination charges which were substantially higher than its rivals.

Table 2: The impact of asymmetric regulation on termination charges (March 2006)[7]

| | Daytime | Evening | Weekends |
|---|---|---|---|
| $O_2$ | 6.4 | 6.3 | 3.1 |
| *Orange* | 7.6 | 5.4 | 4.3 |
| *T-Mobile* | 8.1 | 4.0 | 4.0 |
| *Vodafone* | 8.5 | 3.4 | 2.8 |
| *H3G* | 15.6 | 10.8 | 2.5 |

The mobile industry in the UK currently consists of four mature and roughly equal sized mobile networks, each with around 15 million subscribers, and a recent entrant (*H3G*) with fewer subscribers:

Table 3: Subscriber numbers in 2001 and 2005[8]

| | *Vodafone* | $O_2$ | *Orange* | *T-Mobile* | *H3G* |
|---|---|---|---|---|---|
| Active subscriber numbers in 2001 (m) | 11.0 | 11.1 | 12.4 | 10.3 | n/a |
| Active subscriber numbers in 2005 (m) | 14.8 | 17.0 | 14.9 | 15.3 | 3.5 |

This compares to a total of around only 8 million subscribers at the time of initial regulation in 1998—see Competition Commission (2003, Figure 6.1). Total annual retail revenue for mobile networks in 2007 was about £13 billion, and mobile call termination generated annual revenue of around £2.5 billion (Ofcom, 2007, p7). As mobile penetration has grown, and as the cost of making off-net MTM calls has fallen (see below), the importance of MTM calls in generating termination revenue has risen. Whereas at the time of the Competition Commission's enquiry in 2002, nearly three quarters of mobile termination revenue was from

[6] See Figure 3.38 in Ofcom (2006a).

[7] Taken from Figure 1 in Ofcom (2006b).

[8] See Figure 3.40 in Ofcom (2006a). The current total number of subscribers exceeds the official UK population. This can be explained by the fact that some people have two or more phone subscriptions; for example, one for business and one for personal use; or because they continue to hold an old subscription which they no longer actively use. This is consistent with evidence from Figure 3.46 in Ofcom (2006a) that 10% of UK households do not have access to a mobile phone, similar to the proportion without access to a fixed-line phone.

FTM traffic, now it is only about one third of the total.[9] We will see in section 4.3 that the volume of FTM traffic relative to MTM traffic will play an important role in determining the equilibrium termination charge.

Importantly, and in contrast to the situation in many countries, all mobile networks in the UK are separately owned from the significant fixed networks. This means potential concerns that arise in other jurisdictions about vertical price squeezes and foreclosure are not likely to be an issue in the UK. In section 5 we will discuss how our analysis applies when one of the mobile networks is integrated and also offers a fixed-line service.

The regulated termination charges in Table 1 were calculated using two kinds of markup over estimates of marginal termination costs. The first markup is designed to tax fixed-line callers to subsidise mobile network use in order to stimulate mobile network expansion. As will be explained further in section 3.3, this is consistent with the positive network externalities generated by additional subscribers. The second markup reflects an intended contribution to a mobile network's fixed and common costs. In section 3.4, we argue that including fixed and common costs in this context may be a flawed policy unless there is significant scope for expansion in the mobile market.

The price caps for FTM and MTM termination have been set equal to each other, as were the actual FTM and MTM termination charges set by networks. This contrasts with our initial analysis in section 3 below, where in the absence of regulation we argue that networks may wish to set lower charges for terminating MTM calls (although often the welfare-maximizing charges are the same for the two kinds of calls.) It is worth noting there was no regulatory constraint that prevented networks setting lower MTM charges.

Table 4: Average call charges, pence per minute/message[10]

| | Off-net MTM calls | On-net MTM calls | FTM calls | Text messages |
|---|---|---|---|---|
| 2001 | 26.2 | 5.9 | 14.4 | 8.1 |
| 2005 | 11.3 | 4.2 | 11.5 | 6.3 |

In Table 4 we give an idea of average per-minute retail prices for on-net and off-net MTM calls, as well as FTM calls and text messages. The decline in off-net MTM and FTM call charges evident here is no doubt partly due to the fall in termination charges documented in Table 1. However, the decline in off-net MTM call charges has been particularly dramatic and this likely reflects that a growing number of call plans include some free off-net MTM calls. Despite the narrowing of the price differentials between off-net and on-net calls, though, the difference remains striking. Due in part to this price differential, Table 5 shows the volumes of off-net and on-net calls to be unbalanced. With equal off-net and on-net charges and four

[9] See Figure 3.46 in Competition Commission (2003) and Ofcom (2007, p7).

[10] Cost of FTM calls is taken from Figure 3.22 in Ofcom (2006a). It is a complicated, and largely arbitrary, task to give precise estimates for the prices of the various types of calls and messages originating on mobile networks. This is because mobile networks each offer a wide variety of tariffs, with different monthly rentals (where applicable) corresponding to different volumes of inclusive call minutes and text messages. Other than those for FTM calls, the numbers in Table 4 are taken from Figure 3.39 in Ofcom (2006a), although the method of calculation is not clear from that document.

roughly symmetric networks, one might expect that off-net traffic would be approximately three times greater than on-net traffic, rather than having a *lower* volume than on-net traffic.

In addition to the price differential documented in Table 4, there are at least two other reasons why call volumes are biased towards on-net calls. First, "closed user groups", i.e., groups of subscribers who predominantly make calls within their own group, may be present. Often, such groups have their network subscription decision made centrally (e.g., by their employer's procurement office) and to a single network. To the extent these groups are widespread, this will boost the share of on-net calls in the market. Second, there may be some substitution between MTM and FTM calls. A mobile subscriber, when she is in the home or office, has a choice between calling another mobile subscriber by means of either her fixed line or mobile phone. In many cases, she will just use the cheaper alternative. With the charges in Table 4, this implies she will often want to make an on-net MTM call if the recipient is on the same mobile network, although less so for off-net MTM calls. This will amplify the bias towards on-net call volumes. We discuss this issue further in section 4.3.

Table 5: Shares of types of mobile calls[11]

| | Off-net MTM | On-net MTM | mobile to fixed |
|---|---|---|---|
| % in 2001 | 14.9 | 31.0 | 54.1 |
| % in 2005 | 25.8 | 34.8 | 39.4 |

Since 2003, Ofcom has determined that the mobile *retail* market was effectively competitive.[12] In contrast, UK regulators have consistently ruled that each mobile network has a monopoly of call termination on its own network. The idea is that there is no practical way for people to call someone on the go without calling the person's mobile phone and having the call terminated by the mobile network to which that person has subscribed. This is not to suggest that there are absolutely no substitution possibilities. One objective of this paper is to explore the extent to which this conclusion remains robust to substitution possibilities, both at the wholesale level on the supply side and at the consumer level on the demand side (see section 4).[13]

# 3 Separate FTM and MTM Termination Charges

The principal purpose of the benchmark model in this section is to contrast the pricing of MTM termination with the pricing of FTM termination when each charge is chosen

[11] Data from Figure 3.50 in Ofcom (2006a). There is a typo in the original document in which the proportion of on-net and off-net calls MTM calls was reversed. We exclude international calls, premium rate services and various "other services". Shares are measured in minutes of calls per active mobile connection.

[12] Oftel, *Mobile Access and Call Origination Services Market: Identification and Analysis of Market and Determination on Market Power* (chapter 3), 4 August 2003.

[13] Another possibility is that subscribers may sign up to multiple mobile networks. This could potentially resolve the bottleneck problem if a subscriber's callers can choose to call him on the network with the lower termination charge. However, despite relatively high off-net MTM and FTM pricing, so far this does not seem to have been a very significant constraint on pricing. According to Ofcom (2007b, para. 3.29), in 2006 only 7% of mobile subscribers have more than one mobile phone, and this was typically for reasons other than pricing, such as to separate business and personal calls.

separately. That is, in this model we assume that firms can freely set different termination charges for FTM and MTM traffic. Provided FTM retail call charges are regulated to be equal to cost, we will see that welfare is maximized when both the FTM and MTM termination charges are equal to cost. Without regulation, we will see that mobile networks in this model will wish to set an excessive FTM termination charge (in fact, the monopoly charge), while they wish to set too low a MTM termination charge. Readers less interested in the details of the modelling can skip to section 3.4 for a discussion of the findings.

To model MTM calls, a standard framework of two-way interconnection between symmetric networks is adopted, based on the model of Laffont, Rey, and Tirole (1998b) and Gans and King (2001) in which two mobile networks labeled $i = 1, 2$ offer differentiated services.[14] The networks are assumed to negotiate an industry-wide MTM termination charge, denoted $a$. Mobile subscribers are assumed to be identical in terms of their demand for calls to other mobile subscribers.[15] With this simplification, if subscriber $j$ faces a per-minute charge $p$ for calling subscriber $k$, $j$ will choose to make (an average of) $q(p)$ minutes of calls to $k$. Thus, each subscriber is equally likely to wish to call any other subscriber. Let $v(p)$ be the consumer surplus associated with the demand function $q(p)$, so that $v'(p) \equiv -q(p)$.

Added to this framework is a model of FTM termination described in Armstrong (2002, section 3.1) and Wright (2002). There is a fixed-line network, from which a demand for FTM calls is generated. As in the UK, we assume this fixed sector is separately owned from the mobile sector. Each mobile network unilaterally chooses a termination charge for completing FTM calls, and network $i$'s FTM termination charge is denoted $A_i$. (Where possible, we use upper case notation for calls from the fixed network and lower case notation for calls from mobile networks.) We assume that in the first stage firms negotiate a reciprocal MTM interconnection price $a$, and subsequently, in a second stage, they set their FTM termination charges $A_i$ together with their retail tariffs to mobile customers.[16]

If the retail price for FTM calls to mobile network $i$ is $P_i$ per minute, suppose that there are $Q(P_i)$ FTM minutes of calls to each subscriber on network $i$.[17] We assume the fixed network can set different call charges to different mobile networks to reflect the networks' different FTM termination charges.[18] In general, we expect the price $P_i$ to be an increasing

[14] Throughout the paper it is assumed the parameters are such that an equilibrium exists. Laffont, Rey, and Tirole (1998b) discuss the conditions for such an equilibrium to exist in the retail pricing stage, which requires that the MTM termination charge does not differ too much from cost and/or that the mobile networks are sufficiently differentiated. More generally, for equilibrium to exist at all stages of the game we require that networks be sufficiently differentiated.

[15] Dessein (2003), Hahn (2004) and Valletti and Houpis (2005) develop models where subscribers differ in their demand for calls.

[16] The results obtained do not change when the FTM termination charge is instead set in stage two, with retail prices set in a subsequent third stage. However, the case with simultaneous setting of the FTM termination charge and retail prices simplifies the exposition.

[17] One might also consider mobile-to-fixed calls. However, in the simple frameworks presented here, these calls play no significant role in the analysis and are ignored. (A mobile network would just set the price for such calls equal to its cost of providing such calls in the models we present.) Similarly, fixed-to-fixed calls play no role in the analysis and are ignored.

[18] An alternative assumption is that the fixed network cannot "price discriminate" in this way, perhaps because callers do not always know which mobile network they are calling. Without price discrimination, the market failures identified in the following analysis are typically amplified. For instance, a small mobile network's termination charge has only an insignificant impact on the fixed network's average cost of providing

function of the FTM termination charge $A_i$, and write $P_i = P(A_i)$. For instance, it may be that

$$P(A) = C + A \ , \tag{1}$$

where $C$ is the fixed network's marginal cost of originating a call. In this case, the FTM call charge is equal to the fixed network's total cost of making such calls. Such pricing could arise as a result of the regulation of the fixed network or competition between fixed networks.[19] (In section 4.3 we argue that substitution between fixed and mobile calls might also induce the FTM charges in (1).) Let $V(\cdot)$ be the consumer surplus function associated with the demand function $Q(\cdot)$, so that $V'(P) \equiv -Q(P)$. Define

$$F(A) \equiv (A - c_T)Q(P(A)) \tag{2}$$

to be a mobile network's profit, per subscriber, from providing termination services for the fixed network when its FTM termination charge is $A$.

Each mobile firm is assumed to incur a marginal cost $c_O$ of originating a call and a marginal cost $c_T$ of terminating a call, so the actual marginal cost of a MTM call is $c_O + c_T$. In addition, there is a fixed cost $f$ of serving each mobile subscriber, which includes the subscriber's handset, billing costs, and so on. For now, assume that FTM and MTM calls are independent markets, and that the call charge in one market does not affect the demand in the other market. Figure 1 depicts our stylized model of the mobile industry.

As depicted in Figure 1, denote firm $i$'s on-net MTM call charge by $p_i$ and its off-net MTM call charge by $\hat{p}_i$. In addition, the firm charges a fixed (rental) charge $r_i$ for subscribing. If firm $i$'s market share is $s_i$, its subscribers make a fraction $s_i$ of their calls on-net and the remaining $1 - s_i$ calls off-net. Then a subscriber's utility if she joins that network is

$$u_i = s_i v\left(p_i\right) + (1 - s_i) v\left(\hat{p}_i\right) - r_i \ . \tag{3}$$

We assume a Hotelling specification for subscriber choice, and the market share of network $i$ given the pair of utilities $\{u_1, u_2\}$ available from the networks is

$$s_i = \frac{1}{2} + \frac{u_i - u_j}{2t} \ . \tag{4}$$

Here, $t$ is a parameter which represents the degree of product differentiation in the market for mobile subscribers. Note that in this benchmark model we assume there is an exogenously fixed number of mobile subscribers, which is normalised to 1. (Appendix A analyzes a more complicated but more realistic model where the number of mobile subscribers increases if networks offer better deals, and this analysis is summarised in section 3.3.)

Given an initial choice for the MTM termination charge $a$, suppose that network $i$ chooses its own charges to be $(p_i, \hat{p}_i, r_i, A_i)$. Then network $i$'s profit is

FTM calls, and so the network has a unilateral incentive to set an extremely high termination charge, since such a charge does not cause demand for calls from the fixed network to its subscribers to fall significantly. Indeed, FTM call charges will be *above* the monopoly level (see Gans and King (2000) and Wright (2002)).

[19] Indeed, in the UK the FTM termination charge and the FTM retail price have declined together over the period 2001-2005 (Ofcom, 2006b, Figure 3.38) although not one-for-one as assumed in (1): recently Ofcom (2007, para. 3.22) concluded about two-thirds of the reductions to termination charges had been passed through to the FTM call charge.

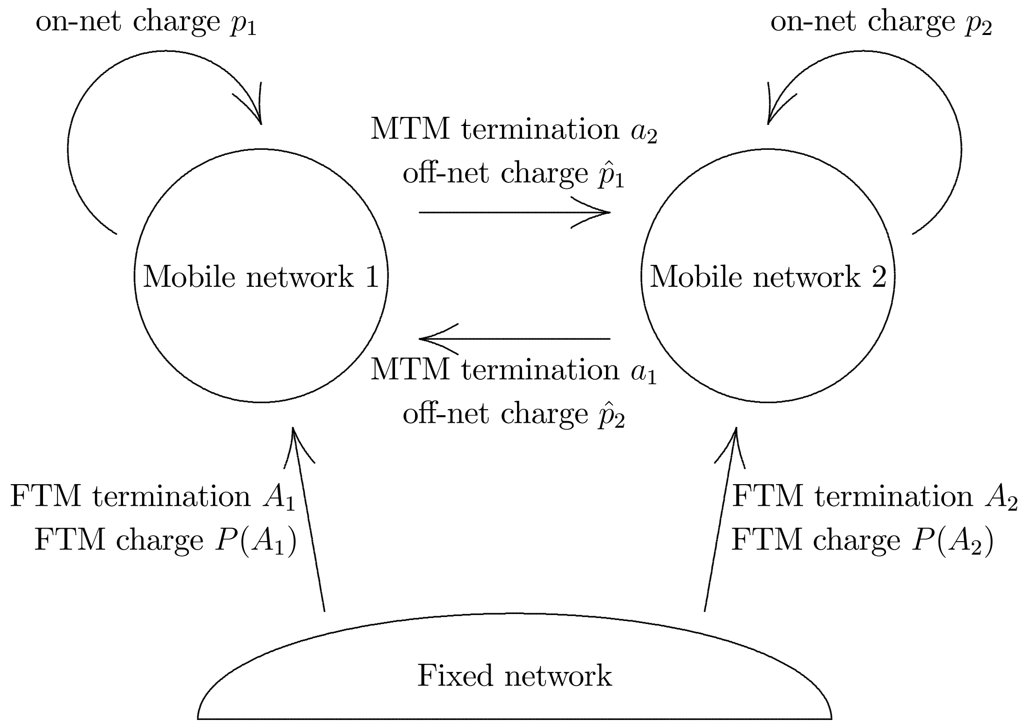


Figure 1: Call Termination on Mobile Networks

$$\pi_i = s_i \times \Bigg[ \underbrace{r_i - f + s_i \left(p_i - c_O - c_T\right) q(p_i) + (1-s_i)(\hat{p}_i - c_O - a) q(\hat{p}_i)}_{\text{retail services}} + \underbrace{(1-s_i)\left(a - c_T\right) q(\hat{p}_j)}_{\text{MTM termination}} + \underbrace{F\left(A_i\right)}_{\text{FTM termination}} \Bigg]. \qquad (5)$$

This consists of the retail profit from supplying service to its subscribers, the profit from providing termination for the rival mobile network, and the profit from providing termination for the fixed network.

## 3.1 Fixed-to-mobile call termination

First, consider network $i$'s incentive to choose its FTM termination charge, $A_i$. The profit expression (5) shows that each network's unregulated FTM termination charge will be chosen to maximise its profits from FTM call termination, $F(\cdot)$. This is a dominant strategy for each network, regardless of choices for retail tariffs and the MTM termination charge. By setting $A_i$ to maximize $F$, each firm will be able to subsidize subscribers to the maximum extent, thereby increasing market share without having to lower profit per subscriber. We denote the resulting unregulated termination charge by $A_M$, to indicate it is the termination charge that would be chosen by a monopoly mobile network.

Welfare (as measured by the sum of consumer surplus and profit) in the FTM segment when the termination charge is $A$ is

$$\underbrace{V(P(A))}_{\text{consumer surplus}} + \underbrace{F(A)}_{\text{mobile termination profit}} + \underbrace{[P(A) - C - A]\, Q(P(A))}_{\text{retail profit by fixed network}} , \tag{6}$$

which simplifies to

$$V(P(A)) + [P(A) - (C + c_T)]\, Q(P(A)) \ .$$

As one would expect, this is maximized by setting the FTM call charge equal to the cost of such calls:

$$P(A_W) = C + c_T \ ,$$

where $A_W$ denotes the welfare-maximising FTM termination charge. This ensures fixed-line callers face a FTM price equal to the true marginal cost of their calls, namely $C + c_T$. When the FTM call charge is equal to the fixed network's cost, so that (1) holds, welfare is maximized with a FTM termination charge equal to cost:

$$A_W = c_T \ .$$

On the other hand, if $P(A) > C + A$ then welfare is maximized by setting $A_W < c_T$ to counteract the markup present in the FTM retail charge.

## 3.2 Mobile-to-mobile call termination

Suppose that the mobile networks each set the FTM termination charge $A$, which could be $A_M$ or any other (perhaps regulated) level of $A$ given that the choice of $a$ will turn out not to depend on $A$ in this benchmark model. First, we derive the equilibrium call charges given $A$ and $a$. In a symmetric equilibrium, each network will share half the market. Therefore, from (3) and (4), firm $i$'s market share is unchanged if it modifies its charges $(p_i, \hat{p}_i, r_i)$ in such a way that

$$\tfrac{1}{2}v(p_i) + \tfrac{1}{2}v(\hat{p}_i) - r_i = u_i \tag{7}$$

is unchanged. From (5) and (7), in a symmetric situation the network's profit in terms of $(p_i, \hat{p}_i)$ is

$$\pi_i = \tfrac{1}{2}\left[\tfrac{1}{2}v(p_i) + \tfrac{1}{2}v(\hat{p}_i) - u_i + \tfrac{1}{2}(p_i - c_O - c_T)q(p_i) + \tfrac{1}{2}(\hat{p}_i - c_O - a)q(\hat{p}_i) + \text{constant}\right] \ .$$

It follows that $p_i$ is chosen to maximize $v(p) + (p - c_O - c_T)q(p)$ and $\hat{p}_i$ is chosen to maximize $v(p) + (p - c_O - a)q(p)$. Therefore, in equilibrium each network will set the on-net call charge $p$ and off-net call charge $\hat{p}$ given by

$$p = c_O + c_T \ ; \ \hat{p} = c_O + a \ , \tag{8}$$

so that call charges are equal to the respective marginal costs of making calls.[20]

[20] This result holds much more generally. It holds allowing for asymmetric market shares and allowing for more than two firms. One case where the result does not apply is when consumers care about receiving calls. Berger (2005) considers this and shows that the on-net call charge is adjusted downwards to reflect the call externality subscribers enjoy from being called more often from others on the same network. Conversely, the off-net call charge is adjusted upwards to reduce the number of calls subscribers receive on rival networks, thereby reducing the rival's ability to compete.

Having determined the equilibrium call charges, we complete the analysis of retail tariff decisions by considering the choice of rental charge, $r$. Analogously to (2), write

$$M(a) \equiv (a - c_T)q(c_O + a) \tag{9}$$

for the profit from MTM termination when the MTM termination charge is $a$. From (5) and (8), network $i$'s profit is

$$\pi_i = s_i \times [r_i - f + (1 - s_i)M(a) + F(A)] \ . \tag{10}$$

Expression (4) implies that firm $i$'s market share $s_i$ satisfies

$$s_i = \frac{1}{2} + \frac{r_j - r_i + (2s_i - 1)(\bar{v} - \hat{v}(a))}{2t} \ ,$$

where $\bar{v} \equiv v(c_O + c_T)$ and $\hat{v}(a) \equiv v(c_O + a)$. Solving this explicitly in terms of $s_i$ implies that

$$s_i = \frac{1}{2} - \frac{r_i - r_j}{2(t - (\bar{v} - \hat{v}(a)))} \ . \tag{11}$$

Finally, substituting (11) into (10), maximizing with respect to $r_i$ and setting $r_i = r_j = r$ shows that the equilibrium rental charge is

$$r = f + t - F(A) - \bar{v} + \hat{v}(a) \ . \tag{12}$$

From (3) and (12), subscriber utility is equal to

$$u = \tfrac{1}{2}\bar{v} + \tfrac{1}{2}\hat{v} - r = F(A) - \tfrac{1}{2}\hat{v}(a) + \tfrac{3}{2}\bar{v} - f - t \ . \tag{13}$$

Clearly, subscriber utility increases with both the termination charges $A$ and $a$ (at least for $A$ up to the monopoly level $A_M$). This observation implies that firms and the regulator can use relatively high termination charges as a means to expand the number of mobile subscribers, as discussed in the next section.

The final variable to determine is the choice of $a$. Substituting (12) into (10) shows that industry profit in the mobile sector is

$$\Pi(a) = \tfrac{1}{2}M(a) + \hat{v}(a) + t - \bar{v} \ . \tag{14}$$

In particular, the FTM termination charge $A$ has no impact on equilibrium profits in the mobile sector. By contrast, mobile networks care about the MTM termination charge. Without regulation, the industry will choose $a$ to maximize (14). The derivative of (14) is

$$\Pi'(a) = \tfrac{1}{2}\left\{(a - c_T)\, q'(c_O + a) - q(c_O + a)\right\} \ ,$$

which is clearly negative when $a = c_T$. In other words, as in Gans and King (2001, Proposition 2), mobile networks prefer a MTM termination charge set below cost. In fact, if demand is concave ($q'' \leq 0$), mobile networks prefer the lowest feasible MTM termination charge. If negative termination charges are not feasible, this implies that networks prefer a so-called "bill-and-keep" arrangement, whereby the MTM termination charge is equal to zero. Since the socially optimal MTM call charge, for both on-net and off-net calls, is equal to the cost $c_O + c_T$, expression (8) implies that the efficient MTM termination charge is $a = c_T$, just as with FTM termination. In particular, optimal regulatory policy treats the two termination charges symmetrically. We deduce that unregulated firms in this model will choose a MTM termination charge which is too low relative to the efficient level, in contrast to incentives concerning FTM termination.

### 3.3 Mobile market expansion

One limitation of the benchmark model just outlined is that the number of mobile subscribers was assumed to be constant and did not change in response to the deals on offer. However, historically at least, the market has instead expanded dramatically (see Table 3). While in a recent survey only 10% of households in the UK market were without access to a mobile phone (Ofcom, 2006a, p157), there may nonetheless remain some marginal subscribers, and subsidizing these subscribers to have a mobile phone could still generate externalities to others.[21] Incorporating market expansion effects can have important implications for the analysis of FTM and MTM termination charges. It will also help us explain why collectively mobile firms resisted regulated reductions in their FTM termination charges, and will play a key role in our analysis of the coordinated choice of a uniform termination charge in section 4.1.

In this section we modify the benchmark model so that number of subscribers network $i$ attracts is $n_i = \phi(u_i, u_j)$ rather than (4), where the total number of mobile subscribers, denoted $N \equiv n_1 + n_2$, is no longer constant. Subscriber utility at firm $i$ is modified from (3) to be

$$u_i = v_0 + n_i v(p_i) + n_j v(\hat{p}_i) - r_i \ . \tag{15}$$

Here, $v_0$ is a subscriber's utility from other mobile services, most notably from calls to the fixed and international network. (This parameter plays no role when the market size is constant and was ignored elsewhere in the paper.) Firm $i$'s profit is modified from (5) to be

$$\pi_i = n_i \times [r_i - f + n_i (p_i - c_O - c_T)\, q(p_i) + n_j(\hat{p}_i - c_O - a) q(\hat{p}_i) + n_j (a - c_T)\, q(\hat{p}_j) + F(A_i)] \ .$$

It is immediate that each firm will set its FTM termination charge $A_i$ to maximise its profits from termination $F(\cdot)$, so $A = A_M$ in equilibrium. The potential for market expansion has no impact on a mobile network's incentive to set a high FTM termination charge, since even with a fixed market size the termination charge was set at its profit-maximizing level.

When the MTM termination charge is $a$, equilibrium call charges still reflect calling costs, so that (8) holds. Therefore, firm $i$'s profit is

$$\pi_i = n_i \left[r_i - f + n_j M(a) + F(A)\right] \ . \tag{16}$$

Incorporating elastic subscriber participation into the benchmark model introduces market-level network effects (as more subscribers join, it becomes more attractive for others to get a mobile phone). In order to make further progress, it is necessary to specify subscriber demand $\phi$ explicitly. To that end, suppose that (4) is modified in the following way: if the two mobile networks offer utilities $u_1$ and $u_2$, then firm $i$ attracts

$$n_i = \frac{1}{2} + \frac{u_i - u_j}{2t} + \lambda u_i \tag{17}$$

subscribers. (This model of consumer demand is sometimes known as the "Hotelling model with hinterlands".) Here, $\lambda \geq 0$ represents the magnitude of the market expansion possibilities. In order to ensure market expansion is non-explosive, $\lambda$ cannot be too large and the

[21] In a recent survey of 621 subscribers, Ofcom (2006c) reports that 12% knew at least one person who had subscribed to be mobile network for the first time in the previous year, and that on average they made 10 calls and received 9 calls per month from these new subscribers.

following condition is assumed:

$$2\lambda\bar{v}\left(1+\lambda t\right)<1\ . \tag{18}$$

Using this model, in Appendix A we establish that the equilibrium market size is increasing with $F$, reflecting that high FTM termination profits will feed through into attractive tariffs for mobile subscribers, which will in turn induce more people to subscribe to mobile networks. This implies mobile profits increase as the common FTM termination charge is increased above cost. The efficient FTM termination charge is also now above cost, since greater subscription will benefit all users (fixed and mobile) since they have more people to call. Nevertheless, the efficient FTM termination charge is still below the unregulated level. Starting from this profit-maximizing charge, a small reduction in the charge has a second-order impact on the profit from call termination, and hence only a second-order impact on the number of mobile subscribers. However, it has a first-order impact on the retail price of FTM calls. Therefore, welfare rises with a reduction in the FTM charge from the unregulated level.[22] In sum, with elastic subscriber participation there remains a rationale for regulatory control of the FTM termination charge, albeit reduced.

What about the MTM termination charge when market expansion is possible? Appendix A shows that firms still jointly prefer to choose a MTM termination charge which is below cost, just as in the benchmark model. However, the efficient MTM termination charge is, like the FTM termination charge, now strictly above cost. The reason is that mobile subscriber surplus is increased by a high MTM termination charge (as in the benchmark model). In sum, if in this extended model networks can feasibly set different termination charges for FTM and MTM traffic, then without regulation networks would choose too high an FTM charge and too low a MTM charge. The socially efficient charges, though, are now both above the cost of supplying termination (and are typically set at different levels, if feasible).

## 3.4 Discussion

**FTM termination:** Despite competition between mobile operators in the retail market, the equilibrium FTM termination charge is equal to the monopoly charge. The result does not depend on the competitiveness of the market for subscribers (it does not depend on product differentiation $t$ and it would not depend on the number of firms if our benchmark model was extended to allow more firms). Thus, it is perfectly possible that one side of the mobile market (the retail market for mobile subscribers) is highly competitive, yet the other side (FTM call termination) is essentially a series of monopolies. By contrast, the socially efficient FTM termination charge is lower. If market expansion possibilities are not significant, the efficient charge is approximately equal to the cost of providing termination; if expansion is possible then the efficient charge will be somewhat above cost in order to stimulate this market expansion, but it is always below the unregulated level.[23]

[22] The finding that the welfare-maximizing FTM termination charge is above cost but below the unregulated level has been noted in the existing literature (Armstrong (2002), Wright (2002), and Valletti and Houpis (2005)), but previous models have unrealistically assumed away MTM calls.

[23] During the 2002 Competition Commission enquiry, *Vodafone* stated that the unregulated charge was in range 17 ppm to 20 ppm. At this time, the cap on $O_2$ and *Vodafone*'s termination charge was 9.3 ppm, so *Vodafone* was suggesting that the FTM termination charge would roughly double without the charge control (Competition Commission, 2003, paras. 2.440–2.445). Ofcom (2007, para. 7.49) estimated more recently

There is no reason why only large (or incumbent) mobile networks should have their FTM termination charges regulated, while small (or new) networks are free to set whatever charge they wish. A small network has the same incentive and ability to set a monopoly FTM termination charge as a large firm. Indeed, if incumbent firms are regulated while new firms are not, new firms have a powerful advantage over incumbent firms, in that they have greater ability to fund subsidies to attract subscribers. This asymmetric treatment of firms in the industry is likely to lead to a distorted pattern of supply. A recent illustration of what happens if some firms are regulated and some are not was presented in Table 2 above, where, on weekdays at least, *H3G* took advantage of its then less regulated position to set termination charges which were roughly double those levied by its more regulated rivals.

In the benchmark model, competition between networks to attract subscribers meant higher FTM termination charges resulted in lower prices for mobile subscribers. Each subscriber brings a profit $F(A)$ and competition acts to pass these profits onto the subscribers themselves. Correspondingly, if regulation squeezes out profits in FTM call termination, this will lead to price rises for mobile subscribers. This result is often termed the "waterbed" effect. In the benchmark model with constant market size (section 3.1) there is a 100% waterbed effect, in that reduced profits from one source are completely clawed back from subscribers so that the overall profit impact is zero. Thus, without market expansion effects, firms should not object to regulatory intervention to bring each firm's FTM termination charge down from the unregulated level to the socially efficient level. Nevertheless, in reality mobile networks *do* object to such a policy. Allowing for market expansion effects of lower tariffs (section 3.3), we found that the resulting waterbed effect is less than 100%. As such, mobile networks will collectively strictly prefer a higher FTM termination charge, and so have an incentive to lobby against proposed regulation to bring down FTM termination charges to efficient levels.

A feature of a strong waterbed effect is that unregulated monopoly profits from FTM call termination are largely passed onto mobile subscribers in the form of low rental charges (e.g., free or subsidized handsets). To the extent this occurs, the market failure associated with FTM termination does not necessarily lead to excessive profits by mobile networks, but rather to a sub-optimal balance of retail prices. Of course, this observation should not affect the welfare analysis: if high margins on FTM call termination lead to negative margins on services to mobile subscribers, there is allocative inefficiency regardless of whether overall profits in the mobile sector are excessive or not. Relatedly, mobile firms "advanced the argument that, because most people had a mobile phone, what they lost in high termination charges they gained in low access and outbound call charges".[24] However, even if all fixed-line subscribers had a mobile phone, this argument is not correct: since high termination charges lead to allocative inefficiency, the total "size of the cake" is shrunk, and the gain from handset subsidies is *smaller* than the losses caused by high FTM call charges.[25] Nevertheless,

that the unregulated FTM termination charge would be about 24 ppm, compared with its current regulated charge of about 5 ppm, and so the regulator believed there would be a five-fold increase if regulation were abandoned.

[24] See Competition Commission (2003, paras. 2.390–2.400).

[25] Ofcom (2007, para. A19.37) estimates that consumers will be better off by £3.2 billion in present value terms as a result of regulation in the period 2007-2011 compared to the unregulated alternative. (Ofcom's analysis assumed that competition was sufficiently strong so that excess profits were eliminated, and so

as we discuss in section 4.3, if FTM calls can substitute for MTM calls, the degree of overlap between the set of fixed-line and mobile subscribers does become relevant.

Recall from section 2 that UK policy has been to allow mobile networks to recover a portion of their fixed costs via a surcharge on FTM termination charges. While such a procedure is fairly standard in setting access prices for a regulated fixed-line network, it is less clear that this policy should be applied to competitive mobile networks. If a strong waterbed effect operates, setting higher FTM termination charges will not provide operators with a significant contribution towards their fixed and common costs, and the policy will simply act as an (inefficient) transfer of surplus between fixed and mobile subscribers. In this environment, setting high termination charges in an attempt to allow for fixed and common cost recovery is likely to be a flawed policy. On the other hand, if market expansion possibilities are significant, the waterbed effect is weaker and above-cost termination charges could, if necessary, be used to help cover the networks' fixed costs (which are not part of our formal model), as well being used to stimulate market expansion.

An implicit assumption in the benchmark model is that fixed-line callers alone determine the number and length of FTM calls. If in practice the volume of FTM calls is jointly determined by the caller and receiver, then high FTM termination charges could lead mobile networks to pay subscribers for receiving FTM calls (rather than merely reducing rental charges), so as to stimulate FTM call volumes. For instance, in 2006 *H3G* announced that it would pay its subscribers 5 ppm for receiving calls, a marketing tactic which surely was motivated by its highly profitable termination charges (see Table 2).[26]

This last point is also helpful in understanding the difference between the "caller-pays" regime as used in the UK, in which unregulated FTM termination charges and call charges are relatively high, and a "receiver-pays" regime used in the United States, whereby the price of a FTM call is low but mobile subscribers often incur charges for receiving calls. In the United States, reciprocity requirements imposed by regulators mean that FTM termination charges are set equal to those for mobile-to-fixed termination, which are quite likely below the cost of mobile termination.[27] With little termination revenue, mobile operators may choose to charge mobile subscribers directly to recover their costs and/or to induce their subscribers to discourage incoming traffic. For instance, with a charge for receiving calls, a subscriber may keep her phone switched off to eliminate incoming calls, something which is often inefficient. The key point is not that the problem of high FTM termination charges is solved by imposing a receiver-pays regime, but rather that a regulatory requirement to set low FTM termination charges may induce networks to charge subscribers for receiving calls.

**MTM termination:** Why do firms wish to set a below-cost MTM termination charge, with or without market expansion possibilities, if free to do so? Unless firms set a low termination charge, call charges (as in (8) above) will be such that it is more expensive to call off-net than on-net. In such a situation, subscribers will, all else equal, prefer to join a larger network

"consumer" and "total welfare" standards coincide.)

[26]There are likely to be inefficiencies involved with the use of reception payments. For instance, some people might call their own mobile phones from office lines to obtain rebates.

[27]One reason for this regulatory approach was the fact that mobile telephone numbers were on the same numbering range as fixed network numbers, posing the obvious problem that callers could not tell they were calling a potentially much more expensive mobile number.

since they can then make a larger fraction of their calls at the cheaper on-net rate. In other words, the market will exhibit (positive) network effects. As is well known, in such markets competition is particularly fierce and profits are low.[28] Firms can overturn this effect by setting a low MTM termination charge, which results in off-net call charges which are below on-net charges. In this case, subscribers will prefer to join a smaller network, which acts to relax competition.

In this framework, mobile subscribers benefit both from high FTM and high MTM termination charges. However, the reason is quite different in the two cases. Mobile subscribers benefit from high FTM charges since there is a waterbed effect at work, and profits from this source are passed (at least partially) onto mobile subscribers. High FTM termination charges are a means of transferring surplus from fixed callers to mobile recipients (and in part to mobile networks if the waterbed effect is not complete). High MTM termination charges act to intensify competition between mobile operators due to strengthened network effects. When the MTM termination charge is raised, this imposes a direct cost on subscribers since they must pay more for off-net calls, but this is outweighed by the lower rental charge they pay. High MTM termination charges have little impact on fixed callers (unless market expansion possibilities are significant), and so a high MTM termination charge acts principally as a means by which to transfer surplus from mobile networks to their subscribers. This also implies that high MTM termination charges cannot be used as a means to cover fixed costs in the industry.

**A linear demand example:** It is useful to illustrate these results with a specific numerical example. Consider a linear demand example, in which costs are normalised to zero ($C = c_O = c_T = f = 0$), where the parameter $v_0$ from the market expansion model is set equal to zero, where the network differentiation parameter is $t = \frac{1}{2}$, and where the two demand functions are $Q = \mu\left(\frac{1}{2} - a\right)$ and $q = (1-\mu)\left(\frac{1}{2} - a\right)$. The parameter $\mu$ measures the relative importance of FTM calls relative to MTM calls. The unregulated FTM termination charge is $A_M = \frac{1}{4}$ (which is broadly in line with Ofcom's (2007b, para. 7.49) estimate of the unregulated FTM termination charge), while in the absence of market expansion possibilities firms will wish to choose as low a MTM termination charge as is feasible. The efficient choice of each termination charge in the absence of network expansion possibilities is $A_W = 0$. As discussed, in the absence of regulation, firms will choose the monopoly FTM termination charge regardless of $\lambda$ and $\mu$, while the choice of the MTM charge will depend on these parameters but will always be below cost (which in turn is weakly lower than the efficient MTM termination charge).

What is the size of the waterbed effect in this example, and how sensitive is mobile penetration to changes in the FTM termination charge? In general, these effects will depend on $\mu$ and the choice of MTM termination charge $a$, but for simplicity consider the case where $\mu = 1$ so that MTM calls are absent. In this case, when $\lambda = 0$ there is no scope for market expansion and as in the benchmark model there is a 100% waterbed effect (network profits do not depend on $A$ in expression (14)). If $\lambda = \frac{1}{4}$, one can calculate using the analysis in Appendix A that mobile industry profits rise by approximately 0.017 when the FTM termination charge is increased from zero to the monopoly charge. The profit due to FTM

[28] See Farrell and Klemperer (2007) for an overview.

termination rises from zero to 0.0537. It follows that around 32% of termination profits are retained by the industry and the waterbed effect is 68%.[29] In the same situation, mobile subscriber numbers increase by 3.1%. Table 6 reports these comparative statics, as well as those for other values of $\lambda$.

Table 6: Effects of increase in FTM termination charge from marginal cost to monopoly level ($\mu = 1$)

| | Increase in subscribers | Size of waterbed effect |
|---|---|---|
| $\lambda = 0$ | 0 | 100% |
| $\lambda = \frac{1}{8}$ | 1.6% | 80% |
| $\lambda = \frac{1}{4}$ | 3.1% | 68% |
| $\lambda = \frac{1}{2}$ | 6.1% | 51% |

The welfare-maximizing pair of termination charges for this linear demand example are given in Table 7. The table illustrates the general analysis which showed that both charges should be set above cost whenever market expansion is possible. In the cases considered, in which the waterbed effect varied between 50% and 80%, the potential for market expansion implies that the efficient regulated FTM termination charge is between one-fifth and two-fifths of the way between marginal cost and the unregulated monopoly charge.[30] We also see that it is generally efficient, if feasible, to set different termination charges for the two kinds of traffic. However, the difference is negligible when market expansion possibilities are small (e.g., $\lambda = \frac{1}{8}$). In this example at least, more significant market expansion possibilities lead to higher optimal FTM and MTM termination charges.

Table 7: Welfare-maximizing choice for non-uniform termination charges[31]

| | $\mu = 0$ | $\mu = \frac{1}{2}$ | $\mu = 1$ |
|---|---|---|---|
| $\lambda = 0$ | $a = 0$ | $a = A = 0$ | $A = 0$ |
| $\lambda = \frac{1}{8}$ | $a = 0.051$ | $a = 0.051,\ A = 0.050$ | $A = 0.050$ |
| $\lambda = \frac{1}{4}$ | $a = 0.072$ | $a = 0.071,\ A = 0.076$ | $A = 0.074$ |
| $\lambda = \frac{1}{2}$ | $a = 0.085$ | $a = 0.084,\ A = 0.102$ | $A = 0.099$ |

[29] Genakos and Valletti (2007) empirically estimate the magnitude of the waterbed effect, and conclude (page 2): "although the waterbed is shown to be high, our analysis also provides evidence that it is not full: accounting measures of profits are positively related to [termination rates]. Mobile firms tend to keep part of termination rents instead of passing them on to their customers, and thus suffer from cuts in termination rates."

[30] Ofcom estimated that the "externality surcharge" needed to reflect the benefits of market expansion was 0.30 pence per minute, or around 5% of their estimated costs of call termination (Ofcom, 2007b, p349). Their chosen methodology, following the earlier analysis of the Competition Commission (2003), was to assume that mobile networks could, at least in part, target a subsidy to new or marginal subscribers rather than to the body of existing subscribers. It is not clear exactly how this could be done in practice, even if mobile networks wanted to do so. Our analysis assumes that the subsidy is applied equally to all mobile subscribers, which means a substantially higher level of subsidy will be needed.

[31] The choice of $A$ is irrelevant when $\mu = 0$ and the choice of $a$ is irrelevant when $\mu = 1$.

**A puzzle:** As mentioned in the introduction, we are not aware of any regulator who has been concerned that firms will set MTM termination charges which are unduly low. Rather, the worry has been that MTM termination charges will be set too high. Related to this, even when given the opportunity to do so, mobile networks do not typically set their MTM termination charges below their FTM charges.[32] These observations seem to contradict the implications of this benchmark model with separate termination charges (with or without market expansion possibilities). In the next section we resolve this puzzle by taking into account constraints that force FTM and MTM termination charges to be locked together. Before turning to this analysis, though, we note some other potential explanations for why mobile networks may in practice prefer high MTM termination charges.

If retail tariffs were simply linear prices, perhaps because a sufficient number of subscribers have pre-paid mobile contracts involving only per-minute charges, it is possible that firms would agree to set above-cost MTM termination charges.[33] Assuming linear pricing, Armstrong (1998), Laffont, Rey, and Tirole (1998a) and Carter and Wright (1999) showed that networks prefer to set MTM termination charges above cost in order to raise the retail price of MTM calls and so undo retail competition. However, these early papers did not allow for differential on-net and off-net retail pricing of MTM calls, which is permitted and common in most countries (see Table 4 for the UK). Allowing for this, Laffont, Rey, and Tirole (1998b) show that firms may prefer below-cost MTM termination charges when there is sufficient competition between mobile operators, even with linear pricing. Firms have two pricing instruments, and so committing to high off-net prices through a high MTM termination charge may not raise profits since it makes firms compete more intensely through their on-net charges. Berger (2004) extends this analysis to allow subscribers to obtain utility from receiving calls, and shows the tendency for networks to prefer a below-cost termination charge is enhanced. Moreover, given that many subscribers do face monthly rental charges or other types of fixed fees, we think the use of nonlinear tariffs adopted in this paper (and by most of the more recent literature) is reasonable. The fact that some subscribers pay linear prices, without on-net and off-net differentials, is not likely to be the principal reason networks have a bias towards high MTM termination charges.

Another reason why incumbent networks might prefer high MTM termination charges is that high charges could act to deter entry or induce exit of a smaller rival. By setting above-cost MTM termination charges, incumbent networks can induce network effects which make entry less attractive for the newcomer. With high MTM termination charges, off-net calls will be more expensive, which particularly hurts a small network since the bulk of its subscribers' calls will be off-net.[34] An additional effect of high off-net call prices will be to

[32] In the UK, in the period 1998 to 2002 the two leading mobile networks' MTM termination charge was not regulated, and in the period since 2003 to the time of writing, mobile networks' MTM termination charges were *capped* (at the same level as the FTM termination charges), but networks were free to set MTM charges below this cap. At no time have networks chosen to set their MTM and FTM termination charges at different levels.

[33] Ofcom (2006a, Figure 3.36) shows that approximately two-thirds of UK mobile subscribers use pre-paid service. However, many pre-paid tariffs still incorporate nonlinear pricing of various kinds, and many involve off-net and on-net price differentials. In addition, subscribers on contracts generate about four times more revenue on average than pre-pay subscribers—see Table 4.39 in Ofcom (2007a).

[34] Calzada and Valletti (2007) demonstrate that incumbent networks can sometimes use high MTM termination charges to deter entry. A related literature examines whether a smaller or entrant network should

reduce the number of calls received by a small network's subscribers, thereby further reducing its ability to compete when call externalities are important.[35] This explanation for high MTM termination charges is complementary to our explanation outlined in the next section. If incumbent networks faced the threat of entry, then the affect of their chosen termination charge on entry would be an additional factor determining the equilibrium charge, and the threat of entry presumably implies that the chosen termination charge will be higher than without this threat.

At least two further explanations have been proposed in the literature for why firms may prefer high MTM termination charges, although both rely on considerably more subtle mechanisms. First, Cherdron (2002) considers a setting where calling patterns are biased towards peers (closed user groups). Setting above cost MTM termination charges is a way to endogenously differentiate the networks, so that consumers prefer to join the network that their peers join, thereby reducing competition between operators. Second, Höffler (2006) argues in a dynamic model with heterogenous users that higher MTM termination charges can allow a collusive outcome to be sustained for a wider range of discount rates. This is because higher termination charges make deviating less attractive given that deviating in his model involves attracting the high-usage consumers and facing a net outflow of calls.

None of these explanations indicate why a mobile network typically sets its FTM and MTM termination charges uniformly. In our view, there is a simpler explanation for why firms prefer high MTM termination charges, which is based on networks not wanting to undercut their high FTM termination charges. We turn to this explanation in the remainder of the paper.

# 4 Uniform Termination Charges

Up until this point, we have assumed that mobile networks negotiate an industry-wide MTM termination charge, which could freely be set at a different level from the FTM charge. However, in practice FTM and MTM charges are not set at different levels. As mentioned in the introduction, one reason why mobile networks may be forced to set a uniform charge is the ability of networks to arbitrage between significant differences in FTM and MTM termination charges.[36] In addition, a network whose FTM termination charge is regulated (or in the shadow of regulation) may be unwilling to choose a lower charge for MTM termination if it suspects the regulator may use that information subsequently to tighten FTM regulation.

In this section we analyse three aspects of the choice of uniform termination charges. First, we isolate the impact of requiring the same termination charge for both kinds of traffic (section 4.1). Here, we assume that networks negotiate their uniform charge. Since without the uniformity constraint we know that firms will set a high FTM charge and a low MTM charge, when the two charges are locked together the uniform charge will be higher than the

be allowed to set a higher MTM termination charge than its larger rivals to help offset its disadvantage in this respect (see Peitz (2005)).

[35] Hoernig (2007) analyses the impact of on-net and off-net price differentials on the profitability of small networks in the presence of call externalities. He shows that larger firms will choose greater differentials than smaller firms.

[36] In our discussions with industry representatives in various countries, this was the explanation given for why mobile operators set a uniform termination charge for both types of calls.

jointly chosen MTM charge. Indeed, often the unregulated jointly-chosen uniform charge will be above the efficient level. Thus, the assumption that charges must be uniform can by itself make sense of the regulatory concern that charges would be too high rather than too low.

Second, in section 4.2 we suppose that the uniform termination charge is chosen unilaterally by individual networks. When each network's FTM and MTM termination charge is constrained to be the same, we believe it is more natural to assume that unregulated uniform termination charges are chosen unilaterally rather than in a coordinated fashion.[37] When charges are chosen unilaterally, it is intuitive (and confirmed in the analysis) that networks will choose a higher charge than when the charge is jointly chosen. The reason is that one network will not internalize the negative impact on a rival's business when it sets a high charge, while this is taken into account when charges are jointly determined. Thus, the assumption that unregulated charges are chosen unilaterally acts to reinforce the tendency towards higher charges already present in section 4.1. Finally, in section 4.3 we consider the impact of demand substitution between FTM and MTM calls (when the uniform termination charge is unilaterally chosen). This acts to *reduce* a firm's incentive to set a high termination charge, although never by enough to make the unregulated charge be inefficiently low.

## 4.1 Coordinated choice of uniform termination charges

In the situation where the number of mobile subscribers is constant, so that a 100% waterbed effect is present, the analysis of section 3.2 regarding the MTM charge applies precisely to the modified situation where the uniform termination charge is collectively set through negotiation. The reason for this is that the level of the FTM termination charge has no impact on the industry's equilibrium profits, and so when firms choose the termination charge collectively, profits from FTM termination have no impact on their overall profits. Therefore, the firms' incentives are exactly as if there is only MTM traffic. In this case, the unregulated uniform termination charge would be set below the efficient level. Thus, the puzzle remains as to why regulators in practice are concerned with termination charges that are set too high.

A more interesting and realistic case is where there is the potential for market expansion, so that the waterbed effect is only partial and firms benefit from high FTM termination charges. To analyse this case we use the model presented in section 3.3, but assume networks jointly choose a uniform termination charge. (Detailed analysis is in Appendix B.)

Firms will agree on a uniform termination charge which lies between the two charges they would choose if FTM and MTM termination could be priced separately. Firms will trade-off the need to protect their FTM termination profits with the worry that a higher uniform termination charge will strengthen network effects and hence the intensity of retail competition. In other words, when both types of termination charges must be set at the same level, MTM termination will act as a constraint on the competitive bottleneck problem that arises with respect to FTM termination. The larger the proportion of MTM calls, the more important is this constraint. On the other hand, the incentive to protect FTM termination

[37] With FTM and MTM termination charges tied together, if networks jointly choose a MTM termination charge they will also effectively be choosing their FTM termination charge jointly. This could raise price-fixing concerns.

profit given the waterbed effect is less than 100% will mean they will set a higher charge than the MTM termination charge they would choose with separate charges.

It is ambiguous whether firms will collectively agree on a uniform termination charge that is above or below cost, or above or below the efficient level. We illustrate with the linear demand example introduced in section 3.4. Table 8 reports the unregulated firms' joint choice of uniform termination charge:

Table 8: Joint profit-maximizing choice of uniform termination charge

| | $\mu = 0$ | $\mu = \frac{1}{2}$ | $\mu = 1$ |
|---|---|---|---|
| $\lambda = 0$ | $a$ as low as feasible | $a$ as low as feasible | $a$ is irrelevant |
| $\lambda = \frac{1}{8}$ | $a = -2.618$ | $a = -0.111$ | $a = 0.250$ |
| $\lambda = \frac{1}{4}$ | $a = -1.465$ | $a = 0.066$ | $a = 0.250$ |
| $\lambda = \frac{1}{2}$ | $a = -0.270$ | $a = 0.153$ | $a = 0.250$ |

Clearly, when $\mu = 0$ the equilibrium termination charge is below cost, since there are no FTM calls and so nothing to counter the firms' desire to relax retail competition by means of negative network effects. Likewise, where market expansion possibilities are limited, there is a very strong waterbed effect and FTM termination profits do not act as much of a brake on the incentive to set low termination charges. However, when $\mu$ and $\lambda$ are not too small, the profit-maximizing uniform charge is above cost, but below the monopoly charge $A_M = \frac{1}{4}$.

The next table shows the corresponding welfare-maximizing charges:

Table 9: Welfare-maximizing choice of uniform termination charge

| | $\mu = 0$ | $\mu = \frac{1}{2}$ | $\mu = 1$ |
|---|---|---|---|
| $\lambda = 0$ | $a = 0$ | $a = 0$ | $a = 0$ |
| $\lambda = \frac{1}{8}$ | $a = 0.051$ | $a = 0.050$ | $a = 0.050$ |
| $\lambda = \frac{1}{4}$ | $a = 0.072$ | $a = 0.074$ | $a = 0.074$ |
| $\lambda = \frac{1}{2}$ | $a = 0.085$ | $a = 0.098$ | $a = 0.099$ |

Here, the efficient termination charge is above cost whenever market expansion is possible (as in section 3.3). When both kinds of traffic are present ($\mu = \frac{1}{2}$), the efficient uniform charge lies between the efficient charges reported in Table 7 for the two different types of calls. (Clearly, the outer columns in Tables 7 and 9 coincide.) Comparing Tables 8 and 9 shows it is ambiguous whether unregulated firms will choose too high or too low a uniform termination charge when they do so in a coordinated fashion. In broad terms, if MTM traffic is particularly significant then the danger is that firms agree to set too low a charge, while if FTM traffic is more significant the danger is that firms choose too high a termination charge. In a range of intermediate cases (e.g., $\mu = \frac{1}{2}$ and $\lambda = \frac{1}{4}$ in the example), the unregulated jointly-chosen termination charge will be reasonably close to the efficient level, in which case requiring networks to coordinate over a uniform FTM and MTM termination charge, if such a policy were feasible, may lead to a reasonably efficient outcome.

## 4.2 Unilateral choice of termination charges

As noted previously, given FTM and MTM termination charges are locked together, in the absence of regulation it is more natural to assume each mobile operator sets its own uniform termination charge. In this section we analyze the situation in which the uniform termination charge is chosen unilaterally by each firm.[38] We use the benchmark model presented in section 3, modified so that (i) each network sets the same charge for FTM and MTM termination and (ii) the two networks can set different uniform termination charges. We initially assume that the number of mobile subscribers is constant in this analysis, and then briefly discuss the extension to allow for market expansion.

Denote network $i$'s uniform termination charge for both FTM and MTM traffic by $a_i$. Then following the same logic as in expression (8), firm $i$'s call charges are

$$p_i = c_O + c_T \ ; \ \hat{p}_i = c_O + a_j \ , \tag{19}$$

and, as in (10), its profit is

$$\pi_i = s_i \times [r_i - f + (1 - s_i)M(a_i) + F(a_i)] \ . \tag{20}$$

From (4), firm $i$'s market share $s_i$ satisfies

$$s_i = \frac{1}{2} + \frac{s_i \bar{v} + (1 - s_i)\hat{v}(a_j) - r_i - [(1 - s_i)\bar{v} + s_i \hat{v}(a_i) - r_j]}{2t}$$

and so

$$s_i = \frac{1}{2} + \frac{r_j - r_i + \frac{1}{2}(\hat{v}(a_j) - \hat{v}(a_i))}{2t + \hat{v}(a_1) + \hat{v}(a_2) - 2\bar{v}} \ . \tag{21}$$

We will ultimately compare the equilibrium termination charge with the "monopoly" charge. Since the MTM demand function $q(\cdot)$ and the FTM demand function $Q(\cdot)$ might differ, we define the monopoly charge $a_M$ to be the common charge which maximizes total termination profit per subscriber; i.e., holding fixed the number of subscribers on each network. Given that there are two networks, half of the MTM calls received by each subscriber are off-net calls. Each subscriber therefore generates termination revenue of $F + \frac{1}{2}M$, and the monopoly termination charge, $a_M$, satisfies

$$F'(a_M) + \tfrac{1}{2}M'(a_M) = 0 \ . \tag{22}$$

In the absence of strong market expansion possibilities, we will show that the equilibrium termination charge will be below $a_M$.

[38] The situation in which MTM termination charges are chosen unilaterally, but not necessarily at the same level as FTM charges, is a special case of this framework in which FTM demand is set to zero. This situation was previously examined by Gans and King (2001), who assumed away FTM calls. However, without the requirement of uniformly set termination charges, it is not clear why networks would set their charges unilaterally. For instance, in the examples we know of where firms have attempted to set different charges for the two kinds of termination (France and New Zealand), firms have set a *low* MTM charge, which could not be an equilibrium with unilaterally-chosen charges. Moreover, even if for some other reason networks set their charges unilaterally, this does not explain why in practice networks set FTM and MTM termination charges at the same level.

As before, we require the networks to be sufficiently differentiated so that network effects are not explosive. From (21), a simple sufficient condition to ensure this is

$$2t > 2\bar{v} - \hat{v}(a_M) \ . \tag{23}$$

This rules out any deviation by network $i$ (even to the point that $\hat{v}(a_i) = 0$), when the other network sets its termination charge no higher than the monopoly level $a_M$, as in the proposed equilibrium found below. Without such a condition, there is a possibility that network $i$ could profitably deviate: by setting a very high termination charge, network $i$ could make off-net calls from the other network prohibitively expensive, so that under certain consumer expectations about which network people will join, everyone will subscribe to network $i$.

Differentiating (20) with respect to $r_i$ and setting equal to zero yields

$$r_i = f - \frac{s_i}{\frac{\partial s_i}{\partial r_i}} - (1-2s_i)M_i - F_i = f + s_i(2t + \hat{v}(a_1) + \hat{v}(a_2) - 2\bar{v}) - (1-2s_i)M_i - F_i \ , \tag{24}$$

where we have written $M_i = M(a_i)$ and $F_i = F(a_i)$, and the second equality follows from (21). Substituting this value for $r_i$ into (20) shows that firm $i$'s profit is

$$\pi_i = s_i^2(2t + \hat{v}(a_1) + \hat{v}(a_2) - 2\bar{v} + M_i) \ . \tag{25}$$

Moreover, substituting $r_i$ as given in (24) and the corresponding expression for $r_j$ into (21) shows that

$$s_i = \frac{1}{2} + \frac{F_i - F_j + \frac{1}{2}(\hat{v}(a_j) - \hat{v}(a_i))}{6t + 3(\hat{v}(a_1) + \hat{v}(a_2)) - 6\bar{v} + 2(M_1 + M_2)} \ , \tag{26}$$

which gives the equilibrium market shares in terms of the two networks' termination charges.

Substituting (26) into (25), differentiating with respect to $a_i$ and imposing symmetry ($a_1 = a_2 = a$) implies the first-order condition for the uniform, unilaterally chosen termination charge $a$ is

$$\left.\frac{\partial \pi_i}{\partial a_i}\right|_{a_i=a_j=a} = \tfrac{1}{4}\left(M'(a) - q(c_O + a)\right)$$

$$+ \left[F'(a) + \tfrac{1}{2}q(c_O + a)\right]\left(\frac{2(t + \hat{v}(a) - \bar{v}) + M(a)}{6(t + \hat{v}(a) - \bar{v}) + 4M(a)}\right) = 0 \ . \tag{27}$$

Evaluating the derivative in (27) at $a = c_T$ yields

$$\left.\frac{\partial \pi_i}{\partial a_i}\right|_{a_i=a_j=c_T} = \tfrac{1}{3}Q\left(C + c_T\right) + \tfrac{1}{6}q\left(c_O + c_T\right) > 0 \ . \tag{28}$$

This demonstrates that the unilaterally chosen termination charge will be above marginal cost. Since the efficient termination charge is $a = c_T$, it follows that unregulated firms will unambiguously choose too high, rather than too low, a uniform termination charge.

This result was previously derived in Gans and King (2001, Proposition 1), in the case where FTM calls were absent.[39] Gans and King (p. 417) attribute this result to a "standard

[39] Setting $Q\left(C + c_T\right) = 0$ in (28) gives Gans and King (2001, Proposition 1).

double marginalisation result that arises when firms set prices for complementary services independently." Our interpretation is different. When network $i$ raises its uniform termination charge there are three effects: (i) its profit from supplying call termination to its mobile rivals and to the fixed network increases (just as with FTM termination in the benchmark model); (ii) its rival is forced to raise its off-net call charge, which directly increases firm $i$'s market share; and (iii) it amplifies the differential pricing between on-net and off-net calls in the market which, as we argued in section 3.4, tends to intensify competition for subscribers. The equilibrium choice of termination charge will trade off these three effects. Effects (i) and (ii) suggest that a firm will want to set its termination charge *above* the monopoly level, since (ii) gives a reason to boost the charge in addition to extracting termination profits. But (iii), which is a strategic effect working through softening retail competition, puts downward pressure on termination charges. In this model expression (28) shows that the net incentive is to set the termination charge above the efficient level. Thus, the two direct effects dominate the incentive to set a low charge in order to relax competition between networks. Expression (28) also shows that the existence of FTM calls increases the firm's incentive to set its mobile termination charge above cost when its MTM and FTM termination charges must be uniform. The existence of FTM calls increases the direct effect in (i), so further raising the incentive to set above-cost termination charges.

We next compare the unregulated charge in (27) with the monopoly charge (22). Equation (27) can be rewritten as

$$\frac{F' + \frac{1}{2}M'}{F' + \frac{1}{2}q} = \frac{t + \hat{v} - \bar{v} + M}{3\,(t + \hat{v} - \bar{v}) + 2M}\,, \tag{29}$$

where the dependence on $a$ in (29) has been suppressed. It is straightforward to check that the left-hand side of (29) is greater than the right-hand side when $a = c_T$. In contrast, at the monopoly termination charge $a_M$ the left-hand side equals zero (by definition), while the right-hand side is positive from (23). Therefore, there is at least one $a \in (c_T, a_M)$ where (29) holds. Moreover, under relatively mild regularity conditions, the left-hand side is decreasing and the right-hand side is increasing in $a$ in the range $(c_T, a_M)$.[40] Therefore, there is then a unique $a \in (c_T, a_M)$ which satisfies (29).

In sum, the equilibrium termination charge is below the monopoly level but above the efficient level. When networks set a uniform termination charge for FTM and MTM traffic, this model predicts unregulated firms will set a uniform charge which is too high relative to the efficient benchmark, and this justifies regulatory concerns that the charge is too high (rather than too low as our benchmark model predicted in section 3). Nevertheless, the fact that the FTM and MTM termination charges are locked together does mitigate a network's incentive to set monopoly termination charges, and the competitive bottleneck result emphasized in section 3 is softened. The key insight here, as with the coordinated setting of termination charges in section 4.1, is that mobile operators will be constrained in their choice of a uniform termination charge by the fact that setting it too high strengthens network effects, making the firms tougher rivals. To avoid this effect, a network will keep its unilaterally-chosen termination charge below the monopoly level.

[40] One can check that the right-hand side is increasing in $a$ over this range provided that $M(\cdot)$ is concave in $a$. The left-hand side is decreasing in $a$ over this range if the stronger condition $M'' < q'$ holds. (A sufficient condition for this second inequality to hold is that $q(\cdot)$ be weakly concave.)

How close is the equilibrium charge to the efficient level? Consider the linear demand example of section 3.4. Condition (23) requires that $t > \frac{7}{64}(1-\mu)$ which is satisfied in this example where $t = \frac{1}{2}$. If $\mu = \frac{1}{2}$ and $\lambda = 0$ then the solution to (29) is $a = 0.230$, which is some 8% below the monopoly charge $a_M = \frac{1}{4}$. One can check that the right-hand side of (29) is decreasing in $t$ when $a > c_T$. Therefore, the unique $a \in (c_T, a_M)$ which solves (29) is increasing in $t$. That is to say, if the retail market is more competitive, in the sense that services are closer substitutes, then the equilibrium uniform termination charge will be closer to the efficient level. This contrasts with our benchmark model in which the two charges were set independently: there, competitive conditions at the retail level played no role in the determination of either charge. However, in our numerical example we find that the competitiveness of the retail market does not seem to play a very significant role in the determination of the equilibrium uniform charge. The change in $a$ with respect to a change in $t$ is greatest when all calls are MTM but even then increasing $t$ from $t = \frac{1}{4}$ to $t = \frac{3}{4}$ increases $a$ only by about two percent.

Another observation is that if the relative importance of FTM traffic declines, in the sense that $F'(\cdot)$ in (29) is reduced over the relevant range for $a$, the equilibrium termination charge falls too. (The left-hand side of (29) increases with $F'$.) Recall from section 2 that the relative importance of FTM traffic has fallen in recent years in the UK, and so this suggests that firms now have a reduced incentive to set very high termination charges compared to the period when regulation was first introduced. The first row of Table 10 shows the significant reduction in equilibrium termination charges induced by a reduction in the proportion of FTM calls in the case without market expansion. When MTM calls are more important, they play a greater role in mitigating a network's incentive to set monopoly termination charges.[41]

Table 10: Unilateral choice of uniform termination charge

| | $\mu = 0$ | $\mu = \frac{1}{2}$ | $\mu = 1$ |
|---|---|---|---|
| $\lambda = 0$ | $a = 0.197$ | $a = 0.230$ | $a = 0.250$ |
| $\lambda = \frac{1}{8}$ | $a = 0.223$ | $a = 0.241$ | $a = 0.250$ |
| $\lambda = \frac{1}{4}$ | $a = 0.237$ | $a = 0.247$ | $a = 0.250$ |
| $\lambda = \frac{1}{2}$ | $a = 0.251$ | $a = 0.251$ | $a = 0.250$ |

How do these results change when subscriber participation is elastic? Appendix C derives the equilibrium conditions for this case. Evaluating these conditions for the linear demand example, Table 10 shows that market expansion possibilities induce firms to choose *higher* termination charges relative to the outcome with a constant market size, presumably because firms take into account the benefits in terms of market expansion of having a higher termination charge. When market expansion effects are strong enough, the chosen charge might even be above the monopoly level. In these (perhaps rather extreme) cases, a firm's desire to choose the monopoly charge for FTM termination actually tempers their incentive to set an even higher termination charge for MTM calls.

[41]For similar reasons, if MTM termination charges are set unilaterally but are not required to be set at the same level as FTM charges, they will also be set lower than in the uniform case (although from (28) still above the efficient level). For instance, with $\mu = \frac{1}{2}$ and $\lambda = 0$, we find the unilateral choice of a separate MTM termination charge is $a = 0.199$ compared to $a = 0.230$ for the uniform case.

Finally, comparing results across different market configurations, we note that the uniform termination charge is higher when it is unilaterally determined (Table 10) compared to when it is set in a coordinated fashion (Table 8). Given it is set unilaterally, the MTM termination charge is also typically higher when it is required to be set uniformly compared to when networks are free to set distinct FTM and MTM termination charges (the exception is when market expansion effects are so strong that the unilaterally set MTM termination charge is above the monopoly level, which arises in our example when $\lambda = \frac{1}{2}$). As a result, at least in this linear demand example, there is still a significant gap between the efficient and unregulated termination charges (Table 9 vs. Table 10), and thus there remains a rationale for intervention.

### 4.3 Substitution between FTM and MTM calls

The previous literature, including this paper up to this point, has largely treated FTM callers and MTM callers as though they were disjoint groups, although this is clearly not the case. Mobile subscribers who have access to a fixed line can choose between the two types of calls. It turns out that this form of substitution has two beneficial effects on the market. First, it weakens a mobile network's incentive to set a high termination charge. Second, it reduces the need to regulate the fixed network's FTM call charges.

To discuss these points, assume as in section 4.2 that networks unilaterally set a uniform charge for terminating FTM and MTM traffic. Suppose there is full penetration of both the fixed and the mobile networks, so everyone potentially has access to both a fixed and a mobile phone to make calls.[42] When a user has access to a fixed phone, we assume that making calls to a mobile subscriber using a mobile phone is a *perfect* substitute for making the call with a fixed-line phone. Thus, a caller will use whichever method is the cheaper. Since people may have different demands for calling when they have access to a fixed-line phone (for example, when they are at home) from when they do not (for example, when they are on-the-go), we allow what we previously termed the FTM demand function $Q(\cdot)$ to denote demand in the former situation and the MTM demand function $q(\cdot)$ to denote demand in the latter situation. (The fact that people only have access to a fixed-line phone a fraction of the time is encompassed by this general specification.) To make the main points as cleanly as possible, suppose that the cost of making calls using the two methods is the same, i.e., $C = c_O$.

Given these assumptions, a subscriber's utility from network $i$ is modified from (3) to be

$$u_i = s_i v(p_i) + (1 - s_i)v(\hat{p}_i) - r_i + s_i V(\min\{P_i, p_i\}) + (1 - s_i) V(\min\{P_j, \hat{p}_i\}) \ . \qquad (30)$$

Here $P_i$ is the charge for making FTM calls to network $i$. The final two terms in (30) are new compared to (3), and represent the utility of a subscriber being able to make calls either with her fixed-line phone or her mobile phone whenever she has a choice. Previously, we ignored the possibility of such substitution and so implicitly assumed subscribers in these situations could only make these calls using their fixed-line phone. In that case, the corresponding

[42] An alternative kind of substitution between fixed and mobile networks may take place at the subscription level rather than the per-call level, so that some people might give up their fixed line altogether and become mobile-only users. We do not consider this possibility here (see Hansen (2006) for an interesting analysis).

utility $s_i V(P_i) + (1 - s_i) V(P_j)$ was independent of the mobile network a subscriber joins, and so did not affect network $i$'s market share and had no bearing on the analysis. Allowing for substitution, though, implies network $i$'s utility in (30) will depend on whether it allows subscribers to call other mobile subscribers more cheaply using its mobile network than they can do using their fixed-line phone.

Suppose that FTM calls satisfy (1), so that regulation or competition forces the fixed network to set its call charges equal to its perceived cost, i.e.

$$P_i = c_O + a_i \ . \tag{31}$$

(We will see shortly that competition with mobile networks for calls will indeed force the fixed network to price in this way.) As with $\bar{v}$ and $\hat{v}$, we define $\bar{V} \equiv (c_O + c_T)$ and $\hat{V}(a) \equiv V(c_O + a)$. Suppose that in the first stage the two networks have chosen their, possibly unequal, termination charges $a_1$ and $a_2$, and are considering their choice of retail tariff. When choosing its call charges, a mobile network must decide whether or not to undercut the prevailing FTM call charge, which will determine which phone a subscriber uses when she has a choice.

Consider the situation in which a subscriber on network $i$ wishes to make a call to a subscriber on the same network, and where the caller has access to her fixed-line phone. If the caller uses the fixed-line phone, she pays $P_i = c_O + a_i$ and enjoys surplus $\hat{V}(a_i)$, while network $i$ obtains termination profit $(a_i - c_T)Q(c_O + a_i)$. Thus, the joint surplus available to the two parties is

$$\hat{V}(a_i) + (a_i - c_T)Q(c_O + a_i) \ .$$

On the other hand, if the network undercuts the FTM call charge by choosing $p_i \leq P_i$, the subscriber enjoys surplus $V(p_i)$, the firm obtains profit $(p_i - c_O - c_T)Q(p_i)$, where this profit now comes from supplying calls rather than call termination, and the available surplus is

$$V(p_i) + (p_i - c_O - c_T)Q(p_i) \ .$$

If $a_i > c_T$, the latter strategy yields the higher joint surplus, and this second joint surplus is maximized by setting $p_i = c_O + c_T$. We deduce that when $a_i \geq c_T$ firm $i$ will set $p_i = c_O + c_T$, with the result that its subscribers will always use their mobile phone to make calls to others on the same network, even when they have a choice of phone. On the other hand, if $a_i < c_T$, total surplus is reduced if the network strictly undercuts the (low) FTM call charge, and the network might as well set $p_i = c_O + c_T$ in order to achieve the maximum total surplus in those situations in which the subscriber cannot use her fixed-line phone.

Consider next the network's off-net call charge, $\hat{p}_i$. If the subscriber uses her fixed-line phone to call someone on the rival network, her surplus is $\hat{V}(a_j)$ and network $i$ makes nothing. If network $i$ undercuts this FTM call charge, so that $\hat{p}_i \leq P_j$, then the surplus of the subscriber is $V(\hat{p}_i)$ and the network's profit is $(\hat{p}_i - c_O - a_j)Q(\hat{p}_i)$. The joint surplus following the latter strategy is maximized by setting $\hat{p}_i = P_j = c_O + a_j$, in which case the joint surplus with the two strategies is identical. Therefore, the firm and subscriber are indifferent between the two strategies. As such, it is strictly optimal for the network to set $\hat{p}_i = c_O + a_j$, since that also maximizes the joint surplus in those situations where the subscriber can only use her mobile phone. In sum, for all choices of termination charges $a_1$ and $a_2$, it is optimal for each network to set its call charges to reflect its cost of making calls, so that (19) holds.

Note next that, in this model with full penetration and perfect substitutes, the fixed network will be forced to set its call charges as in (31) above, even if these charges are not regulated. Given the mobile networks are setting their call charges as in (19), if the fixed network increases its price above the price in (31) consumers will always use their mobile phone, while if it decreases its price it will offer its service at a loss. Thus, an important feature of demand substitution is that it eliminates the need to regulate FTM call charges.

In situations where $a_i \geq c_T$, it follows that subscriber utility in (30) becomes[43]

$$u_i = s_i \left(\bar{v} + \bar{V}\right) + (1 - s_i) \left(\hat{v}\left(a_j\right) + \hat{V}\left(a_j\right)\right) - r_i \ ,$$

so that from (4) firm $i$'s market share satisfies

$$s_i = \frac{1}{2} + \frac{r_j - r_i + \frac{1}{2}\left(\hat{v}(a_j) + \hat{V}(a_j) - \hat{v}(a_i) - \hat{V}(a_i)\right)}{2t + \hat{v}(a_1) + \hat{V}(a_1) + \hat{v}(a_2) + \hat{V}(a_2) - 2(\bar{v} + \bar{V})} \ .$$

In particular, relative to section 4.2, network effects are more pronounced, and condition (23) now needs to be tightened to $2t > 2\left(\bar{v} + \bar{V}\right) - \hat{v}(a_M) - \hat{V}\left(a_M\right)$.

Network $i$'s profit when $a_i \geq c_T$ is

$$\pi_i = s_i \times \left[r_i - f + (1 - s_i)\left(M(a_i) + F\left(a_i\right)\right)\right].$$

To understand this expression, notice that (i) network $i$'s subscribers will never use a fixed-line phone to call others on the same network, (ii) when the network's subscribers call someone on the other network, they are indifferent about using their mobile or fixed-line phone (when they have a choice), and the network makes no profit from these calls in either case, and (iii) when a subscriber on the rival network calls a network $i$ subscriber and has the option to use their fixed-line phone to make the call, they are indifferent about which type of phone to use, and network $i$ makes the same profit $F(a_i)$ in either event (and where they do not have this option they will use their mobile phone and network $i$ makes profit $M(a_i)$).

Following the same steps as used in section 4.2 shows that instead of (27) the unilaterally-chosen uniform termination charge here satisfies

$$\begin{aligned}\left.\frac{\partial \pi_i}{\partial a_i}\right|_{a_i = a_j = a} &= \tfrac{1}{4}\left(M' + F' - q - Q\right) \\ &+ \left[\frac{q + Q}{2}\right]\left(\frac{2(t + \hat{v} + \hat{V} - \bar{v} - \bar{V}) + M + F}{6(t + \hat{v} + \hat{V} - \bar{v} - \bar{V}) + 4\left(M + F\right)}\right) = 0 \ ,\end{aligned}$$

which can be rewritten as

$$\frac{F' + M'}{Q + q} = \frac{t + \hat{v} + \hat{V} - \bar{v} - \bar{V} + F + M}{3(t + \hat{v} + \hat{V} - \bar{v} - \bar{V}) + 2(F + M)} \ . \tag{32}$$

[43] It can be checked that firm $i$ will not wish to set $a_i < c_T$. Such a policy would induce subscribers to replace on-net calls with FTM calls whenever possible, for which firm $i$ will incur a termination loss. Moreover, even ignoring FTM calls, the analysis in section 4.2 shows that networks will always wish to set an above-cost termination charge when it is set unilaterally.

We are interested in whether demand substitution increases or reduces a network's incentive to set an excessive termination charge, i.e., whether the solution to (32) is above or below that corresponding to expression (29). The comparison is most transparent if we specialize the demand for calls somewhat. Suppose that the underlying demand for calls, which we denote $X(\cdot)$, is the same whether the caller is at home or on-the-go, and that each caller is at home a fraction $\mu$ of the time. In this case, $Q = \mu X$ and $q = (1-\mu)X$, which gives one interpretation of the different weights on the two types of demand in the example given in section 3.4. The equilibrium termination charge in (32) does not depend on $\mu$, since all that matters in (32) is the *sum* of demands $Q + q$, which always equals $X$. With this demand specification, expression (29) coincides with (32) when $\mu = 0$, i.e., when there are no FTM calls. Then one can check that the left-hand side of (29) is increasing in $\mu$ in this example, while the right-hand side is decreasing with $\mu$. It follows that the equilibrium termination charge which solves (29) is increasing with $\mu$. We deduce that the charge $a$ which solves (32) is below the charge $a$ which solves (29). Moreover, the larger $\mu$ is, the greater is the difference between the two.

Thus, we can conclude that demand-side substitution between FTM and MTM calls lowers the equilibrium termination charge towards marginal cost. For instance, using the linear demand example of section 3.4 and assuming that $\mu = \frac{1}{2}$ and $\lambda = 0$ one can show that the equilibrium termination charge which solves (32) is approximately $a = 0.197$, which is some 22% below the monopoly charge (rather than the 8% which applied when there was no such substitution).

In sum, the fact that consumers substitute on-net MTM calls for FTM calls whenever the former's price is lower mitigates a mobile network's incentive to set high mobile termination charges but it does not eliminate it altogether. The intuitive reason for why this substitution reduces a network's incentive to set high termination charges stems from two sources. First, the volume of FTM termination traffic falls since all calls made to people on the same mobile network are made with a mobile phone, even when a fixed-line phone is available. As discussed in section 4.2, when the volume of FTM traffic falls, this will reduce the equilibrium termination charge since firms put more weight on avoiding the intense competition caused by network effects induced by high termination charges. Second, for a given termination charge, network effects are more important when substitution is possible. This is because calls from home (those with demand function $Q$) also have on-net and off-net charge differentials now. This reinforces the first effect, since firms place still more weight on the danger of intensifying competition via high termination charges.

This discussion has assumed that the costs of making the two types of call are equal. It may be more realistic to assume that FTM calls are more efficient than MTM calls when callers have the choice of phone, so that $C < c_O$. However, unless the cost differential is very large, the previous analysis remains valid. A network still has an incentive to set an above-cost termination charge, and it still wishes to under-cut the FTM call charge by setting lower on-net MTM call charges. That is to say, a mobile firm will encourage its subscribers to use the less efficient MTM mode of communication whenever subscribers have a choice. For instance, with the call charges in Table 3 for 2001, it is plausible that mobile subscribers were making too many on-net calls relative to FTM calls. In other words, high mobile termination charges distort competition between FTM and on-net MTM calls, and this danger provides an additional benefit of regulating the termination charge to equal cost.

Finally, when there is the possibility of FTM and MTM call substitution, high FTM call charges and the availability of cheap on-net MTM calls will give people an additional reason to subscribe to a mobile network, which is to avoid high FTM prices for on-net calls. Allowing for elastic subscriber demand, this introduces another avenue by which high mobile termination charges can induce marginal subscribers to join and therefore another positive externality of above-cost termination charges (in addition to the externalities discussed in section 3.3).

# 5 Vertical Integration

The analysis to this point has assumed, in line with the UK market, that mobile networks and fixed networks are separately owned. In this section we briefly discuss the impact of vertical integration in the context of our model. Thus, suppose now that there are two mobile networks as before, but that one of these networks (say, network 1) is integrated with the (monopoly) fixed network.

If the monopoly fixed network is tightly regulated, it turns out that vertical integration has little impact on our previous analysis. Suppose that the fixed network's termination charge for *receiving* calls from mobile subscribers is regulated to be equal to its cost of supplying termination, so that the fixed network obtains no profit from providing this service. Second, suppose that the fixed network's FTM call charges are regulated so that (1) is satisfied. (In the case where the MTM and FTM termination charges can be set differently, network 1's FTM charge $A_1$ is merely an "accounting" price which is used to determine the integrated firm's regulated price for FTM calls to network 1.) Finally, suppose that the level of the fixed network's FTM call charge does not affect other aspects of the fixed network's business (its demand for its other services, its number of subscribers, or the prices it is able to charge for other services).

This collection of assumptions together imply that the fixed network's profit is unaffected by decisions made by its mobile sub-division. As such, our previous analysis applies to this situation where one mobile network is integrated with the fixed network. For example, in the case where termination charges need not be uniform (as in section 3), the integrated mobile network still has an incentive to set its FTM termination charge at the monopoly level, since that maximizes the profit obtained from making FTM calls to that network. Thus, vertical integration has no impact when the fixed network is tightly regulated.

More generally, when the integrated network is tightly constrained in what it can extract from fixed subscribers, it may prefer, where possible, to shift profit to the mobile sector where it is unregulated in the retail market. In this case, the integrated firm's incentives are likely to be similar to a vertically separated mobile firm's incentives as modelled in this paper. On the other hand, if the fixed network is not regulated at all in the charges it can set, or only in an indirect way (e.g., its retail charges are subject to an average price cap), the fixed network's profits will be negatively related to the level of FTM termination charges it has to pay to both mobile networks. In this case, the integrated firm has an incentive to set less distorted FTM prices for calls within its network (so to extract more profit from its fixed subscribers), and also to compete more aggressively against its mobile rival (so that its fixed subscribers need to make fewer expensive calls to the rival network).

These results can be established in the extreme case in which the integrated firm is an unregulated monopolist in the fixed market and only FTM calls are considered. In such a setting, the integrated firm will prefer to maximize the surplus it can extract from its fixed subscribers by pricing FTM calls at cost (perceived cost in the case the calls are off-net) and extracting its subscribers' surplus through the fixed charge. To achieve this, the integrated firm's (internal) FTM termination charge is set at the efficient level so that the firm can set efficient prices for FTM calls within its combined network. In contrast, the mobile-only network will continue to set the monopoly FTM termination charge so as to maximize the subsidy it can offer its subscribers. As a result, the integrated firm will want to compete more aggressively for mobile subscribers. This is to reduce the rival's market share, and hence the share of calls on which its fixed subscribers must pay monopoly prices. This strategy allows the integrated firm to charge a higher fixed charge to its fixed subscribers. If, instead, the rival's termination charge was regulated to be equal to cost, then competition between the integrated firm and the mobile-only firm would not be distorted.

Nevertheless, we are not aware of examples where an integrated firm has lobbied for, or has voluntarily set, a low termination charge. Thus, it may be that situations where the fixed network is tightly regulated—where our analysis with vertical separation applies—is often the more relevant one. A full analysis of the impact of vertical integration on incentives to set termination charges, as well as the incentive to integrate in the first place, will be sensitive to the details of how the fixed network is regulated, and we leave this important issue for future research.

# 6 Conclusions

This paper has aimed to provide a unifying framework for discussing policy towards mobile call termination. Initially, in section 3 we assumed that the charges for terminating FTM and MTM calls could freely be set at different levels. In this setting, a competitive bottleneck exists for FTM termination, so FTM charges are set too high (at the monopoly level), while reciprocal MTM termination charges are set too low so as to relax retail competition between mobile networks. This model accurately reflects UK experience with respect to high FTM termination charges. However, for MTM termination, the model incorrectly predicts that authorities should be concerned that unregulated charges will be too low.

Section 4 resolves this puzzle by noting that wholesale arbitrage implies that a mobile network cannot sustain a FTM termination charge significantly above its MTM termination charge. Taking this supply-side substitution into account, we provide a new analysis in which each network sets a uniform termination charge. To avoid intensifying retail competition through network effects, mobile networks choose their uniform termination charge to be below the monopoly level (at least when market expansion is not a major factor) but above the low level that they would set if MTM termination could be priced separately. The dichotomy that existed in the previous literature is resolved, with unregulated termination charges lying between the two earlier extremes. We establish this result both when networks set their uniform charges unilaterally as well as when they are determined collectively. In the former case, which we think is more reasonable in this setting, the equilibrium termination charge is always above cost. (In the latter case, inefficiently high termination charges arise whenever FTM calls are sufficiently important.)

Section 4.3 shows that demand-side substitution, in which mobile subscribers use their mobile phones to avoid expensive FTM calls, strengthens the constraints imposed by network effects on unregulated termination charges. Assuming a uniform termination charge, such substitution only arises with respect to on-net, not off-net, MTM calls. On-net MTM calls are made by a network's own subscribers, so by setting a low price for on-net MTM calls, a network can insulate such callers from the effects of its choice of high termination charges. As a result, a network's FTM termination profit becomes less important, and the intensification of retail competition through network effects becomes more important, in the choice of termination charges. Thus, with both supply-side arbitrage and demand-side substitution, we find the incentive to set mobile termination charges above cost is mitigated, although not eliminated altogether.

A quite different approach to overcoming the bottleneck present with mobile termination is to try to adjust the bargaining position of fixed-line and mobile operators so that mobile operators no longer hold all the bargaining power.[44] Key to this (without relying on the threat of regulation) would be to relax the fixed-line operator's obligation to interconnect with any mobile operator at the latter's chosen termination charge. For instance, if there were no such obligation, then with a single fixed network and multiple mobile operators, the natural outcome is that the fixed network holds most of the bargaining power. The mobile networks might then compete in a winner-take-all fashion for the right to deliver the fixed network's FTM calls (which they will want to do since their subscribers value receiving calls), resulting in low mobile termination charges. How this might work when off-net MTM calls are also taken into account (for instance, taking into account the supply-side arbitrage and demand side substitution that we consider) remains an open question. It is worth noting, however, that for such a market-based mechanism to replace a regulatory approach would require authorities be willing to let a breakdown in interconnection occur. It is clear in the UK they are not willing to do so at the present time (Ofcom, 2007b, paras. 5.148-5.162).

In the coming years, regulation of the termination of voice calls in mobile networks may also be expanded to include other services. For instance, the EU has introduced regulation of wholesale termination charges for international roaming. Similarly, Ofcom has announced that it plans to review the market for SMS termination (Ofcom, *Wholesale SMS Termination Market Review*, 13 September 2006).[45] Some of the issues discussed in this paper also apply to text messages, which are an increasingly important part of the mobile market. However, networks are better able to set SMS termination charges separately from prevailing call termination charges, and so the analysis in section 3.2 might be more relevant and networks may choose to set low termination charges. On the other hand, networks might wish to set higher SMS termination charges due to a fear that text messages could substitute for calls, in which case some of the issues in section 4.3 could come into play. In future work it would be worthwhile to investigate these issues in more depth.

We close by noting our paper also has implications for the internet, where a debate has emerged over what is termed "net neutrality". This debate has arisen, in part, due to state-

[44] Binmore and Harbord (2005) consider the bargaining process between a large fixed network and a small mobile entrant.

[45] SMS termination charges are already controlled in France and Israel, for instance.

ments of Internet Service Providers (ISPs) in the United States that they intend to start charging content providers for access to their customers. If they do so, the competitive bottleneck story in this paper may become relevant. Consumers (who seek content) typically only sign up to one ISP. However, content providers typically want to reach all consumers regardless of the consumer's chosen ISP. This means ISPs will not compete for content providers with respect to providing access to their customers. The result will be a distorted price structure, with content providers charged too much and too little content being supplied. As with mobile termination, regulations limiting how much ISPs can charge content providers or other networks to access their customers might then need to be considered.

# References


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

——— (2002): "The Theory of Access Pricing and Interconnection," in *Handbook of Telecommunications Economics, Volume 1*, ed. by M. Cave, S. Majumdar, and I. Vogelsang. North-Holland, Amsterdam.

——— (2006): "Competition in Two-Sided Markets," *Rand Journal of Economics*, 37(3), 668–691.

Armstrong, M., and J. Wright (2007): "Two-Sided Markets, Competitive Bottlenecks and Exclusive Contracts," *Economic Theory*, 32(2), 353–380.

Berger, U. (2004): "Access Charges in the Presence of Call Externalities," *Contributions to Economic Analysis and Policy*, 3(1), article 21.

——— (2005): "Bill-and-Keep vs. Cost-Based Access Pricing Revisited," *Economics Letters*, 86(1), 107–112.

Binmore, K., and D. Harbord (2005): "Bargaining Over Fixed-To-Mobile Termination Rates: Countervailing Buyer Power as a Constraint on Monopoly Power," *Journal of Competition Law and Economics*, 1(3), 449–472.

Calzada, J., and T. Valletti (2007): "Network Competition and Entry Deterrence," *Economic Journal*, forthcoming.

Carter, M., and J. Wright (1999): "Interconnection in Network Industries," *Review of Industrial Organization*, 14(1), 1–25.

Cherdron, M. (2002): "Interconnection, Termination-Based Price Discrimination, and Network Competition in a Mature Telecommunications Market," mimeo, Mannheim.

Competition-Commission (2003): *Vodafone, O2, Orange and T-Mobile: Reports on References Under Section 13 of Telecommunications Act 1984 on Charges Made by Vodafone, Orange, O2 and T-Mobile for Terminating Calls Made by Fixed and Mobile Networks.* HMSO, London.

DESSEIN, W. (2003): "Network Competition in Nonlinear Pricing," *Rand Journal of Economics*, 34(4), 593–611.

FARRELL, J., AND P. KLEMPERER (2007): "Coordination and Lock-in: Competition with Switching Costs and Network Effects," in *Handbook of Industrial Organization, Volume 3*, ed. by M. Armstrong, and R. Porter. North-Holland, Amsterdam.

GANS, J., AND S. KING (2000): "Mobile Network Competition, Customer Ignorance and Fixed-to-Mobile Call Prices," *Information Economics and Policy*, 12(4), 301–328.

——— (2001): "Using 'Bill and Keep' Interconnect Agreements to Soften Network Competition," *Economics Letters*, 71(3), 413–420.

GENAKOS, C., AND T. VALLETTI (2007): "Testing the 'Waterbed' effect in Mobile Telephony," mimeo.

HAHN, J.-H. (2004): "Network Competition and Interconnection with Heterogeneous Subscribers," *International Journal of Industrial Organization*, 22(1), 611–631.

HANSEN, B. (2006): "Termination Rates and Fixed Mobile Substitution," mimeo, Norwegian School of Management.

HÖFFLER, F. (2006): "Mobile Termination and Collusion, Revisited," mimeo, Bonn.

HOERNIG, S. (2007): "On-net and Off-net Pricing on Asymmetric Telecommunications Networks," *Information Economics and Policy*, 19(2), 171–188.

LAFFONT, J.-J., P. REY, AND J. TIROLE (1998a): "Network Competition: I. Overview and Nondiscriminatory Pricing," *Rand Journal of Economics*, 29(1), 1–37.

——— (1998b): "Network Competition: II. Price Discrimination," *Rand Journal of Economics*, 29(1), 38–56.

MMC (1999): *Cellnet and Vodafone: A Report on a Reference Under Section 13 of Telecommunications Act 1984 on the Charges Made by Cellnet and Vodafone for Terminating Calls from Fixed-Line Networks*. HMSO, London.

OFCOM (2006a): *The Communications Market 2006*. Office of Communications, London.

——— (2006b): *Mobile Call Termination: Market Review*. Office of Communications, London.

——— (2006c): *Mobile Call Termination: Report of Market Research Findings*. Office of Communications, London.

——— (2007a): *The Communications Market 2007*. Office of Communications, London.

——— (2007b): *Mobile Call Termination: Statement*. Office of Communications, London.

PEITZ, M. (2005): "Asymmetric Access Price Regulation in Telecommunications Markets," *European Economic Review*, 49(2), 341–358.

Valletti, T., and G. Houpis (2005): "Mobile Termination: What is the 'Right' Charge?," *Journal of Regulatory Economics*, 28(3), 235–258.

Wright, J. (2002): "Access Pricing under Competition: An Application to Cellular Networks," *Journal of Industrial Economics*, 50(3), 289–316.

# Technical Appendix: Elastic Subscriber Participation

## A Non-uniform FTM and MTM termination charges

The model setup is as described in section 3.3.[46] Combining (15) with (17) shows that

$$N = \frac{1+\lambda(2v_0 - r_1 - r_2)}{1-\lambda(\bar{v}+\hat{v}(a))} \; ; \; n_i = \frac{N}{2} + \frac{\frac{1}{2}(1+\lambda t)(r_j - r_i)}{t-(1+\lambda t)(\bar{v}-\hat{v}(a))} \; , \tag{33}$$

where $\hat{v}(a) \equiv v(c_O + a)$ as before. (From (18), the denominators in (33) are positive so long as $a$ is not too far away from cost $c_T$.) Write

$$\gamma(a) \equiv -\frac{\partial n_i}{\partial r_i} = \frac{1+2t\lambda - 2\lambda(1+\lambda t)\bar{v}}{D} \; ; \; \mu(a) \equiv \frac{\partial n_j}{\partial r_i} = \frac{1-2\lambda(1+\lambda t)\hat{v}(a)}{D} \; , \tag{34}$$

where

$$D = 2[1-\lambda(\bar{v}+\hat{v}(a))][t-(1+\lambda t)(\bar{v}-\hat{v}(a))] \; ,$$

for the own and cross-price effects on subscriber numbers with respect to changes in the rental charge. Here $\gamma$ is positive, $\mu$ is smaller than $\gamma$, and $\mu$ is positive when $a$ is close to $c_T$ and (18) holds.

From (16) the equilibrium rental charge is given by

$$r = f - \tfrac{1}{2}NM - F + \tfrac{1}{2}N\Lambda \; , \tag{35}$$

where

$$\Lambda(a) \equiv \frac{1+\mu(a)M(a)}{\gamma(a)} \; .$$

Substituting the value for $r$ in (35) into the formula for profit in (16) shows that industry profit is

$$\Pi = \tfrac{1}{2}N^2\Lambda \; , \tag{36}$$

while substituting the value for $r$ in (35) into expression (33) shows that the market size given $a$ and $A$ is

$$N = \frac{1+2\lambda(F+v_0-f)}{1-\lambda(M+\bar{v}+\hat{v}-\Lambda)} \; . \tag{37}$$

[46] The mobile-to-mobile part of the following analysis is somewhat related to section 5 of Dessein (2003), although he does not allow for off-net/on-net call charge differentials. Like us, he shows that the unregulated MTM termination charge is below cost and the efficient MTM termination charge is above cost.

(From (18), the denominator in (37) is positive for $a$ close to $c_T$.) Since $N$ in (37) is increasing with $F$, it follows from (36) that mobile profits are higher with a higher FTM termination profit, $F(A)$, and the waterbed effect present in section 3.1 is no longer complete.

The impact of the MTM termination charge on profit is more complex, since $a$ affects the terms $\gamma$ and $\mu$ in $\Lambda$. For simplicity, consider the impact of a small change in $a$ away from marginal cost $c_T$. One can show that

$$\Lambda'(c_T) = -q(c_O + c_T)\left[\frac{1 - 2\lambda\bar{v}(1 + \lambda t)}{1 + 2t\lambda - 2\lambda\bar{v}(1 + \lambda t)}\right] < 0 \ ,$$

where the inequality follows from (18). From (37) we see that

$$N_a \equiv \left.\frac{\partial N}{\partial a}\right|_{a=c_T} = -\frac{\lambda N \Lambda'}{1 + \lambda\bar{\Lambda} - 2\lambda\bar{v}} > 0 \ ,$$

where $\bar{\Lambda} = \Lambda(c_T)$, and so raising the MTM termination charge above cost induces network expansion. From (36) we have

$$\left.\frac{\partial \Pi}{\partial a}\right|_{a=c_T} = NN_a\Lambda + \tfrac{1}{2}N^2\Lambda' = N^2\Lambda'\left[\frac{1}{2} - \frac{\lambda\bar{\Lambda}}{1 + \lambda\bar{\Lambda} - 2\lambda\bar{v}}\right] < 0 \ ,$$

where the inequality follows (after some manipulation) from (18). Therefore, mobile networks wish to set a below-cost MTM termination charge in order to relax competition for subscribers, just as in the benchmark model without market expansion.

What are the socially efficient termination charges in this framework? When subscriber utility is $u_1 = u_2 = u$, aggregate consumer surplus of mobile subscribers is $\Phi(u) = \lambda u^2 + u$, where $u = v_0 + \frac{1}{2}N(\bar{v} + \hat{v}) - r$. (So $\Phi'(u) \equiv N$.) In equilibrium, $r$ is given by (35). The consumer surplus of fixed-line subscribers from calls to mobile subscribers is $N\hat{V}$, where $\hat{V} = V(P(A))$ is the consumer surplus on the fixed network for each mobile subscriber. Assume that FTM calls are charged at cost, so that expression (1) holds.[47] Then total welfare is obtained by summing mobile sector profit, mobile subscriber surplus and fixed caller surplus, which is

$$W = N\hat{V} + \tfrac{1}{2}N^2\Lambda + \Phi(v_0 + \tfrac{1}{2}N(\bar{v} + \hat{v}) - (f - \tfrac{1}{2}NM - F + \tfrac{1}{2}N\Lambda)) \ . \tag{38}$$

When there are no market expansion possibilities (i.e., $\lambda = 0$), it is optimal to set both the FTM and MTM termination charges equal to cost $c_T$ as in our benchmark model. When $\lambda > 0$ it is socially optimal to set both termination charges above cost. To see this for the FTM termination charge, differentiate (38) with respect to $A$ to obtain

$$\begin{aligned} \left.\frac{\partial W}{\partial A}\right|_{A=a=c_T} &= -NQ + N_A\bar{V} + N\Lambda N_A + N\left\{N_A\bar{v} - \tfrac{1}{2}\Lambda N_A + Q\right\} \\ &= N_A[\bar{V} + N\bar{v} + \tfrac{1}{2}N\Lambda] > 0 \ , \end{aligned} \tag{39}$$

[47] Assume also that the mobile-to-fixed termination charge is regulated to be equal to the fixed network's cost. Otherwise, the extra profits on the fixed network caused by terminating more traffic when the mobile market expands would need to be considered too.

where $\bar{V} \equiv V(C+c_T)$ and $N_A > 0$ is the derivative of $N$ in (37) with respect to $A$ when $A = a = c_T$. Thus, raising the FTM termination charge above cost induces network expansion, and this benefits both fixed-line callers and existing mobile subscribers (and also the mobile networks).

To see the impact of raising the MTM termination charge, differentiate (38) to obtain:

$$\begin{aligned}\left.\frac{\partial W}{\partial a}\right|_{A=a=c_T} &= N_a\bar{V} + [NN_a\Lambda + \tfrac{1}{2}N^2\Lambda'] + N\left\{N_a\bar{v} - \tfrac{1}{2}\Lambda N_a - \tfrac{1}{2}N\Lambda'\right\} \\ &= N_a[\bar{V} + N\bar{v} + \tfrac{1}{2}N\Lambda] > 0 \ ,\end{aligned}$$

which is the same form as (39). Thus, setting the MTM termination charge above cost benefits both mobile and fixed-line subscribers and overall welfare, but it harms the mobile networks.

# B Jointly-chosen uniform termination charge

Now suppose the FTM and MTM termination charges must be equal and firms coordinate on the choice of this uniform charge, denoted $a$. Then the analysis of Appendix A continues to apply except that differentiating (37) with respect to the uniform charge $a$ now yields

$$N' \equiv \left.\frac{dN}{da}\right|_{a=c_T} = \frac{2\lambda Q(C+c_T) - \lambda N\Lambda'}{1+\lambda\bar{\Lambda} - 2\lambda\bar{v}} > 0 \ ,$$

and

$$\left.\frac{d\Pi}{da}\right|_{a=c_T} = NN'\Lambda + \tfrac{1}{2}N^2\Lambda' = N\Lambda\frac{2\lambda Q(C+c_T)}{1+\lambda\bar{\Lambda}-2\lambda\bar{v}} + N^2\Lambda'\left[\frac{1}{2} - \frac{\lambda\bar{\Lambda}}{1+\lambda\bar{\Lambda}-2\lambda\bar{v}}\right] \ .$$

The sign of this derivative is ambiguous. If $Q$ is sufficiently large it is positive and firms will agree to set a uniform termination charge above cost.

# C Unilateral choice of uniform termination charge

Finally, suppose that the two networks choose their uniform termination charge unilaterally. Firm $i$'s profit with charge $a_i$ is modified from (16) to be

$$\pi_i = n_i\left[r_i - f + n_jM(a_i) + F(a_i)\right] \ .$$

Firm $i$ has $n_i$ subscribers, where this is modified from (33) to be

$$n_i = \frac{1}{\tilde{D}}\left\{(1+\lambda(2v_0 - r_1 - r_2))(t - (1+\lambda t)(\bar{v} - \hat{v}(a_j))) + (1 - \lambda(\bar{v} + \hat{v}(a_j)))(1+\lambda t)(r_j - r_i)\right\} \ ,$$

where

$$\begin{aligned}\tilde{D} &= 2\left(1 - \lambda\left(\bar{v} + \frac{\hat{v}(a_1)+\hat{v}(a_2)}{2}\right)\right)\left(t - (1+\lambda t)\left(\bar{v} - \frac{\hat{v}(a_1)+\hat{v}(a_2)}{2}\right)\right) \\ &\qquad + 2\lambda(1+\lambda t)\left(\frac{\hat{v}(a_1)-\hat{v}(a_2)}{2}\right)^2 .\end{aligned}$$

Then the own-price and cross-price effects on subscriber numbers with respect to changes in firm $i$'s rental charge are

$$\gamma \equiv -\frac{\partial n_i}{\partial r_i} = \frac{1 + 2t\lambda - 2\lambda(1+\lambda t)\bar{v}}{\tilde{D}} \; ; \; \mu_i \equiv \frac{\partial n_j}{\partial r_i} = \frac{1 - 2\lambda(1+\lambda t)\hat{v}(a_i)}{\tilde{D}} \; .$$

The rental charge $r_i$ satisfies

$$0 = \frac{\partial \pi_i}{\partial r_i} = -\gamma[r_i - f + n_j M_i + F_i] + n_i[1 + \mu_i M_i]$$

and so

$$\pi_i = n_i^2 \Lambda_i \tag{40}$$

where

$$\Lambda_i \equiv \frac{1 + \mu_i M_i}{\gamma} \; .$$

Expression (40) gives firm $i$'s profit as a (complicated) function of the pair of termination charges $a_1$ and $a_2$, and this can be differentiated with respect to $a_i$ to find the symmetric equilibrium charge when the charges are chosen unilaterally.