Facility based competition in telecommunications

Three essays on two-way access and one essay on three-way access

by

Bjørn Hansen

A dissertation submitted to BI Norwegian School of Management
for the degree of Dr. Oecon

Series of Dissertations 5/2006

BI Norwegian School of Management
Department of Economics

Bjørn Hansen:

Facility based competition in telecommunications
– Three essays on two-way access and one essay on three-way access

Series of Dissertations 5/2006

ISBN: 82 7042 751 9
ISSN: 1502-2099

BI Norwegian School of Management
N-0442 Oslo
Phone: +47 4641 0000
www.bi.no

Printing: Nordberg

The dissertation may be ordered from our website www.bi.no
(Research - Research Publications)

Abstract

In order to reap all benefits from telecommunications, competing firms typically have to cooperate in order to exploit economies of scale and scope. Thus, firms being active in the same market are supposed to compete in some dimensions and cooperate in other dimensions. There is potentially a trade-off between cooperation and competition. In this dissertation four cases of interplay between competition and cooperation are investigated and we find that in some cases there is indeed a trade-off. In some cases (but not all cases) firms can arrange their cooperation such that they are able to soften competition and increase prices. Whether such effects are present or not depends on technology and market characteristics. It is accordingly necessary to carry out case by case analysis in order to assess the interplay between cooperation and competition. A common feature of the four papers in the dissertation is that they take as a starting point a concrete and policy relevant issue where telecommunications firms have to cooperate. Game theoretic models are adapted to each case and particular care is taken in capturing relevant market and technology features.

Acknowledgements

A number of people have contributed to this dissertation at various stages in the process. I would, in particular, like to express my gratitude to my supervisor, Professor Christian Riis. He has given very helpful comments and suggestions to the work in the dissertation.

My former colleague at Telenor R&D, Øystein Foros is co author in two of the papers in the dissertation. To work with Øystein has been inspiring and I'm still working on following his example in effectively finalising papers and submitting them for publication. I have also appreciated the cooperation with Jan Yngve Sand in co-authoring a paper.

In my research I have benefited from the technical insight and understanding of telecommunications provided by colleagues at Telenor R&D, in particular Vegard Masdal. Furthermore, my research has benefited from discussions with Robert Pettersen at NPT on telecommunications regulation.

I would like to thank Telenor, and in particular the R&D department, for financing this dissertation. I am grateful to Professor Øystein Fjeldstad and the Centre for Management of Telecom and Network Industry at the Norwegian School of Management for offering me an opportunity to finalise my work on a PhD that started out at the University of Oslo and continued at Telenor R&D.

Finally, I would like to thank my family – my incredibly patient wife Kjersti as well as my three beautiful girls Mari, Johanne and Kathrine, always willing to put my work in the right perspective.

Oslo, 29 May 2006

Bjørn Hansen

Contents

Section 1

Introduction

Facility based competition in telecommunications
– Three essays on two-way access and one essay on three-way access*

1. Introduction

Deregulation and fast technological change have resulted in a rapid transformation of the telecommunications industry. The initial steps of the deregulatory process were, in most countries, designed so that newcomers invested in some parts of the production chain and then relied on access to other parts of the production chain from a regulated incumbent. The terms and conditions for such access have typically been regulated. The rationale for regulating access is that the regulated segments are considered as bottlenecks, i.e. it is prohibitively costly to duplicate already installed capacity. The local loop in the fixed network is a classical example. The copper cable connecting residential customers to the network will typically have sufficient capacity to carry all telephony and Internet related traffic. Thus it will be socially wasteful if newcomers had to install new cables to reach the household. This kind of access problems is one-way in the sense that newcomers need access but incumbents do not.

One-way access problems have been discussed in the literature, at least since the US Supreme Court's 1912 Terminal Railroad decision (considered as the origination of the essential facilities doctrine). Introductions and overviews of this literature and its applications to telecommunications can e.g. be found in Laffont and Tirole (2000) and Armstrong (2002). The current thesis is not focusing on this kind of access problems.

Telecommunications is characterised by economies of scale on both the supply and the demand side. On the supply side technologies like 3rd generation mobile systems (3G) and fibre optics are characterised by considerable economies of scale. There are accordingly potential gains from making networks cooperate on the supply side. On the demand side the economies of scale are due to network effects. Willingness to pay for network membership typically increases with the number of communication partners. Thus there are potential social gains from making sure that networks interconnect in order to facilitate communication across networks.

* I thank Christian Riis, Øystein Foros and Robert Pettersen for valuable comments to earlier drafts of the introductory chapter.

These access problems are two-way in the sense that each network controls an asset, customers and/or capacity, which is valuable to the other party.

Since there are gains from cooperation, one should expect that networks, being free to negotiate contracts, would be able to design efficient contracts. However, networks are at the same time supposed to compete for the same customers. Thus there is a danger that networks design their access agreements in a way that softens competition. Furthermore, due to the network effects, if networks do not interconnect, one may experience "network tipping", i.e. that all consumers join one of the two networks. Then a network will succeed in gaining a monopoly position. Thus a firm may deny two-way access in an attempt to foreclose the market. At the outset we can accordingly expect that under facilities based competition, unregulated two-way access will in some cases yield efficient outcomes, in other cases there is too much access, and in yet other cases too little. It is accordingly not evident that regulatory intervention is required, and furthermore, if such intervention is required, it is likely that the regulatory design should depend upon characteristics of the market under consideration.

In this thesis we consider four different two-way access situations and we demonstrate that the need for regulatory intervention indeed depends upon technology and market characteristics. The four access problems considered in the thesis are as follows: 1) Telephony interconnection between a competitive mobile sector and a regulated fixed monopoly. 2) Telephony interconnection between competing networks when production costs on the two networks differ. 3) Interconnection between competing Internet service providers, and finally 4) Mutual access to capacity in third generation mobile networks.

The four papers in the thesis confirm that the incentive to provide access, as well as the need for regulatory intervention, indeed depends upon factors like cost asymmetries, technological characteristics and the types of contracts being signed by the players. A prerequisite for policy relevant modelling of network competition is accordingly a fairly detailed knowledge of technology and market characteristics. Therefore, in section 3 of this introduction, we go into some details in describing technology, as well as market outcomes.[1] Prior to the descriptive section on market experience, we

[1] Focusing on one-way access, Faulhaber (2003) argues, based on US experience, that the success or failure of introducing competition by regulatory intervention is explained by whether at least one of two conditions are fulfilled. The first condition being that the dominating firm is excluded from operating in the segment supposed to be competitive and the second being that the interface between the regulated and competitive segments is technically simple, easy to monitor and require little

provide a brief literature overview in section 2. This serves as a context for the subsequent descriptive section. In section 4 we consider some modelling issues that are of importance when considering network competition. In section 5 we provide an overview of the four essays in the thesis. Finally, in section 6 we provide some concluding remarks.

2. A brief introduction to the literature

All papers in the thesis contain a section reviewing relevant literature for that particular paper. Here we will therefore only provide a brief introduction.

Symmetric interconnection of competing telecommunications networks was analyzed by Laffont Rey and Tirole (1998a,b), as well as Armstrong (1998). The type of interconnection studied in these papers is illustrated below:

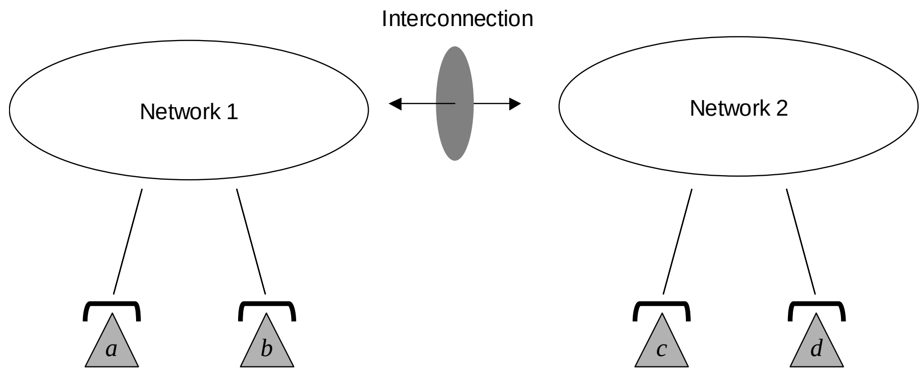

Figure 1, Interconnection

Without interconnection consumers in network 1, respectively 2, can only make calls to other consumers in the same network; on-net calls. Due to interconnection consumers in network 1 can call consumers in network 2 and *vice versa*; off-net calls. An off-net call from say consumer *a* to consumer *d* in the illustration above can be divided into two parts. The first part is origination, i.e. to convey the call from the caller, consumer *a* to the interconnection interface. The second part is termination, to carry the call from the interconnection interface to the receiver, consumer *d*.

In the literature one is typically studying a two stage game where the competing firms, at stage 1, determine the price of termination and then the two networks compete in attracting consumers in the second stage of the

information. Thus, fairly detailed technical knowledge is required by regulators when designing policy.

game. A major insight from the work by Laffont Rey and Tirole (1998a,b) and Gans and King (2001) is that the incentives with respect to setting the price of termination depend upon the contracts in the downstream market. When there is uniform pricing in the downstream market, the mobile firms can use the termination rate as an instrument to soften competition. By raising each other's marginal cost they reduce the competitive pressure. In Laffont Rey and Tirole (1998b) it is demonstrated that this effect changes if one considers two-part tariffs in the downstream market. Then the profits of the mobile firms are independent of the termination rate. Finally, Gans and King (2001) consider network based discrimination, i.e. that the prices for on- and off-net traffic are allowed to differ, and they find that a low termination rate may be used as an instrument to increase profits.

The basic models described above have been extended in a number of ways. Carter and Wright (2003) analyse interconnection of vertically differentiated networks, Dessein (2003) considers consumer heterogeneity and Jeon, Laffont and Tirole (2004) analyse the implications of willingness to pay for receiving calls. Asymmetric termination rates and entry were analysed by Peitz (2005).

In communication networks the utility of being a member of a particular network typically increases with the network size. This is called network effects and was first analysed by Rohlfs (1974). Rohlfs focuses on the existence of multiple equilibria and the resulting problems of starting up a new communications service. The work by Rohlfs does not however take into account that there may be competing firms offering the network service. This is in contrast to Katz and Shapiro (1985) where the implications of network externalities within a competitive environment is analysed. Katz and Shapiro pay particular attention to the choice of compatibility. They find that a large (dominant) firm will prefer too little compatibility, a small firm prefers too much compatibility and the industry would jointly prefer too little compatibility.[2] The seminal analysis by Katz and Shapiro has been extended in a number of ways in the literature. A literature overview can e.g. be found in Liebowitz and Margolis (2002).

3. Market experience

In this section we will discuss some relevant two-way access problems in telecommunications. We will comment on technological characteristics as

[2] This last result is in contrast to a result in one of the papers in the present thesis; Competition and compatibility among Internet Service Providers. This difference is explained by different assumptions regarding downstream competition.

well as market outcomes with respect to two-way access. Finally, we will also briefly discuss the degree of regulatory intervention behind the outcomes.

The implication of the discussion in the current section is that the tension between cooperation and competition is not a transitory phenomenon during a deregulatory process. Even under full facility-based competition there is a potential gain from interconnecting networks. Thus interconnection is an issue for "old" as well as "new" services.

In the literature on two-way access the focus is typically on shared market equilibrium. Parameter-restrictions are imposed in order make sure that this outcome is achieved. Given these restrictions firms will typically gain from providing access. If, however, the parameter restrictions are violated, or some firms expect them to be violated, then the market will tend to tip in one or the other direction. Thus there will be competition for the market instead of competition in the market. Firms competing for the market will not necessarily enter voluntarily into two-way access agreements. Below we will argue that market observations indeed lend support to this assertion.

3.1. Fixed and mobile telephony

Interfaces for interconnecting traditional fixed telephony as well as mobile telephony are well established and all firms providing these services are typically directly or indirectly interconnected to all other networks. This ubiquitous interconnection is the result of a long historical process. Currently it is common to make ubiquitous interconnection a requirement in the license for telephone companies.

There are some examples of non interconnected phone companies from the early days of telephony history. In the period 1881 to 1886 there were two competing phone companies in the capital of Norway, Kristiania.[3] The two companies rolled out parallel access networks and they competed head to head. It was not possible to make calls from one network to the other. The period characterised by access competition ended in 1886 because the local authorities forced the two companies to merge by denying the companies licenses to install new cables until they merged. According to Rinde, (2005), p. 146 it was in particular the merchant community of the city that wanted the two firms to merge. The arguments were twofold; they wanted all phones to be interconnected and they wanted to avoid duplication of civil works from network roll out. Similarly, in the period between the end of the Bell

[3] The name of the city was changed to Oslo in 1924.

patents in 1893 and the Kingsbury commitment in 1913, a number of phone companies independent of the Bell system were denied interconnection (Brock, 2002). During this period, phone companies competed head to head in the US and network effects were used strategically.

During this early period there was a notable difference in the speed of telephony adoption in areas with local competition as compared to areas characterized by local monopolies. As an example, according to Rinde (2005), the rate of telephony adoption in Kristiania was twice as high as the rate of adoption in Copenhagen, which had a local monopoly.

At present, the termination rates of incumbent fixed operators are subject to regulation in both Europe and the US. These rates are supposed to reflect underlying costs. In the US, the Telecommunications act of 96 requires reciprocity of termination rates (see e.g. DeGraba 2004). Thus, fixed-line newcomers are required to charge the same termination rate as the incumbent. This is in contrast to Europe where reciprocity is not embedded in the regulatory framework, thus the national regulators have to decide on whether termination rates should be reciprocal or not.

Taking Norway as an example, only the fixed incumbent (Telenor) is subject to price regulation. Other fixed operators are free to set the termination rate they want.[4] As illustrated below, all the other fixed operators have termination rates above the level charged by the incumbent:

[4] They are mandated to interconnect and they are mandated to have a reference offer (see NPT 2005a, pp 70 - 71).

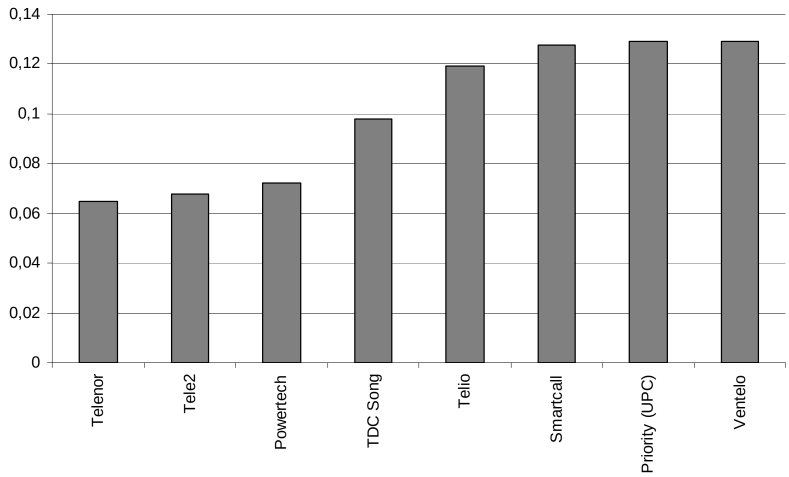

Figure 2, Termination rates fixed networks, (NOK/min) Norway, January 2006, (2 min call, within local termination area)

The variation in termination rates is to some degree correlated with market shares. The smallest networks have the highest termination charges.[5]

Similarly there is considerable variation in termination rates in the Norwegian mobile sector as well, where the two large networks; Telenor and Netcom are subject to differentiated price regulation on mobile termination.

---

[5] There are however some exceptions to this pattern, in particular Ventelo, which is the third largest network with a 5% market share.

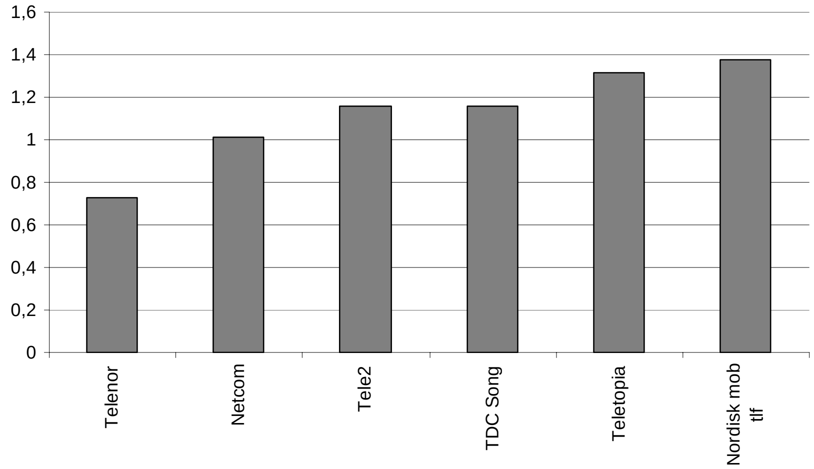

Figure 3, Termination rates mobile networks, (NOK/min) Norway, January 2006, (2 min call)

Teletopia and Nordisk mobiltelefoni are network operators with very small market shares, Tele2 and TDC – Song are virtual operators renting capacity from Telenor Mobil. Mobile termination rates in Norway are currently roughly 10 times higher than the termination rate on fixed networks.

In Sweden the national regulator has committed to introduce reciprocal termination rates between all mobile networks. The regulated reciprocal mobile termination rate in Sweden is below the lowest mobile termination rate in Norway. This is in contrast to Denmark where mobile termination rates are currently not regulated. The Danish mobile termination rates are reciprocal and relatively high (above 1 DKR/min). The mobile market in Denmark is characterised by aggressive competition, low prices and low profits for the mobile firms. The experience from Denmark therefore lends some support to the assertion that it is reciprocity rather than the level of the termination rates that stimulates competition.

3.2. Interconnection in the internet

Interconnection on the internet is arranged quite differently from telephony; it is a hierarchy. The hierarchy and the associated interconnection contracts have evolved as a result of market forces and are not the result of regulatory intervention. The Internet is accordingly an interesting case for comparison.

The Internet is a set of interconnected data networks all using the same system of addresses and protocols enabling communication between users on the different networks. Interconnection is evidently a key element in this architecture. There are two main types of interconnection arrangements in the Internet; peering and transit (see Kende 2002). The Internet hierarchy is divided into three levels or tiers. Each level is characterised by the types of interconnection agreements they are engaged in.

Peering is a barter arrangement where two networks mutually agree to exchange traffic free of charge. The traffic being exchanged is between customers (of customers) on the two peering networks. Peering networks do not accept traffic to third parties (traffic from a peer to other peers). The other type of contract is transit where one network is paid to accept any traffic to and from its contract partner, i.e. also to third parties. With the terminology introduced earlier in this chapter, peering can be characterised as two-way access and transit is one-way access.

At tier 1 of the hierarchy we find the global Internet backbones, such networks are only engaged in peering arrangements. At tier 2 we find networks with a mix of contracts, both peering (typically regionally) and buying transit from one or more of the tier 1 networks. Finally a tier 3 network is not engaged in peering. An overview of the peering arrangements various networks are engaged in can be found at http://www.peeringdb.com/. There is unregulated, seemingly well functioning, competition at each level in this hierarchy. Local access is however an exception. Local loop unbundling as well as other measures are used by regulators to facilitate competition at this level too. However, in 1998 two tier 1 networks, MCI and WorldCom merged. The merger would result in a significant increase in market concentration among tier 1 networks. Both European and US regulators approved the merger under the condition that the Internet business of MCI was divested (see FCC 1998, Cremer et al. 2000 and Economides 2005).

3.3. Internet-based applications

The Internet enables ubiquitous data connectivity. Thus any pair of users can in principle communicate, but they need interoperable applications to facilitate this communication. E-mail is an example of an application (or service) running over the internet such that any e-mail user can communicate with any other e-mail user. This is in contrast to other communications services provided over the internet where interconnection is an issue.

The necessary architecture for providing messenger[6] services and voice over the Internet (VoIP) has some important similarities. In both cases *servers*[7]

[6] A messenger platform enables users to engage in text-based real time dialogues over the Internet.

contain databases linking user names (or phone numbers) to IP addresses such that a user who is logged on can be reached irrespective of physical location. When a communication session is initiated the servers feed address information to the necessary systems such that the actual media stream (e.g. the voice call) is not passing through the server.

Technically, direct interconnect between networks requires servers to be able to "talk to each other" in order to exchange address information. Furthermore, end systems have to be sufficiently compatible (e.g. that the technology for transferring voice to IP packets and back are interoperable).

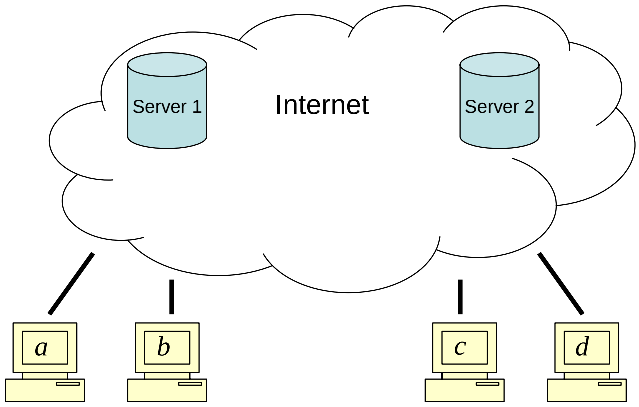

Figure 4, Real time communication on the internet

In the figure above we have illustrated two telephony networks on the Internet. User *a* and *b* as well as server 1 belong to network 1. When user *a* makes a call to user *b*, the software on the originating computer will communicate with the call-server in order to obtain the necessary address information. Provided with this information, the software on computer *a* establishes direct contact with the software on computer *b* and the actual call takes place. The call itself does not pass through the server.

As compared to traditional telephony, the entry barriers for providing services over the internet are relatively low. A newcomer wanting to offer e.g. VoIP must establish a call-server, and distribute necessary software to end users. Thus local access is no longer a bottleneck. Referring to the illustration above, users on network 1 and 2 can already communicate, e.g. by e-mail. The problem is however finding the address of the one you want

[7] For instant messaging and telephony the servers are called IM servers and call-servers respectively. The server functionality may be physically distributed, but logically it works as a database.

to communicate with. Thus the servers must exchange information in order to facilitate interconnection. With the introduction of VoIP a possible new bottleneck is accordingly access to these databases. In addition technical compatibility can be used strategically to gain competitive advantage. The examples provided below demonstrate that denying interconnection of databases as well as incompatible technical solutions indeed are an issue in these markets.

The four large global messenger networks are: MSN Messenger, Yahoo! Messenger, AIM and ICQ. Two of the networks are owned by AOL; AIM and ICQ. These two networks are interconnected. This is in contrast to the other networks. At the time of writing, neither MSN nor Yahoo! offers interconnect to other messenger networks. A user of MSN can accordingly not communicate with a user of ICQ etc. In 1999 Microsoft did try to establish interconnection between MSN Messenger and AIM. The attempts were blocked by AOL which still has a dominant position in this market. Recently it has been announced that MSN and Yahoo! are going to be interconnected during the first half of 2006.[8] Interconnection is evidently a strategic issue for these firms.

Telephony networks based on VoIP are rapidly gaining market shares. Roughly 5% of Norwegian households were connected to a VoIP network as of 1 January 2005. This number is expected to rise to 20% during 2006. The situation in the VoIP market has similarities to the messenger market, and most VoIP networks are accordingly not directly interconnected.[9] Some VoIP networks like Skype have managed to enter the telephony market without interconnecting with other networks. Other VoIP networks are taking a more traditional route by installing a gateway to the established circuit switched telephony networks (a relatively successful example in Norway is Telio).[10] By doing so the entire installed base of telephony users on both fixed and mobile becomes available from VoIP. Since many VoIP networks have a gateway to the traditional telephony network, they are also indirectly interconnected. Thus instead of routing a call between VoIP customers on different networks directly over the Internet, the call is routed

[8] See http://www.microsoft.com/presspass/press/2005/oct05/10-12MSNYahooMessengerPR.mspx

[9] There are some notable exceptions, in particular the US-based network Free world dialup. According to their website, their customers can use FWD to talk with people who use other networks to make calls over the internet.

[10] In addition to the components illustrated in figure 4, interconnection with the circuit switched network also requires a gateway.

via gateways and through the traditional telephony networks. This is illustrated by the dotted line below.

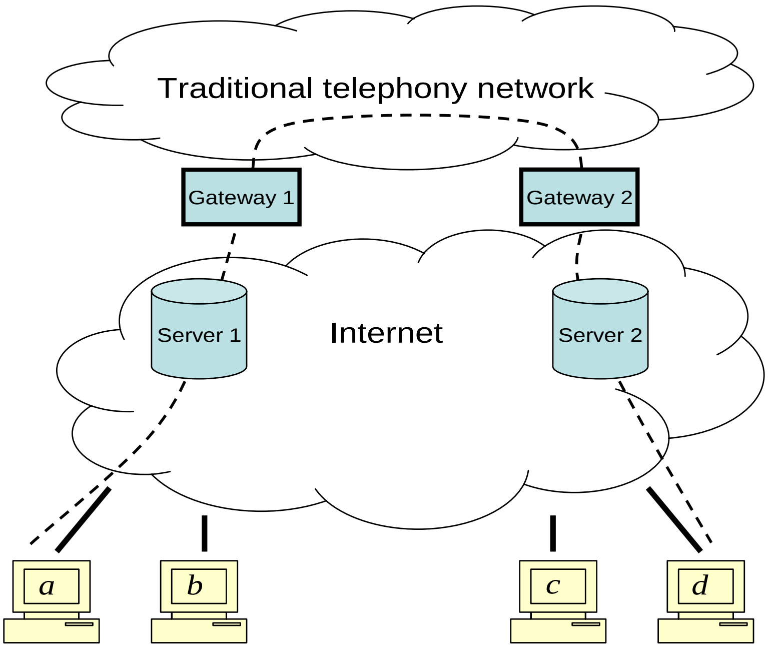

Figure 5, A call routed via the traditional telephony network

This way of facilitating interconnection between different VoIP networks seems inefficient. By interconnecting call-servers and making software sufficiently compatible the call could instead be routed directly over the Internet. As the proportion of customers on VoIP increases, and thus the proportion of VoIP to VoIP calls increases, the significance of the inefficiency will increase.

There are therefore likely to be gains from direct interconnection. A possible future development is then that some firms will deny all kinds of direct interconnection, others will want to interconnect with all others, while a third group of networks will choose targeted degradation. It is accordingly not unlikely that interconnection will continue to be an issue for regulators. The focus will however shift from local access towards access to address information (i.e. interconnection of servers) as well as compatibility and standardization.

3.4. Infrastructure sharing in mobile, national roaming

In most industrialised countries, mobile firms are upgrading their networks to 3G.[11] The cost of obtaining a given geographical coverage is much higher for 3G as compared to 2G. Thus the cost of introducing 3G is significant. Competing networks are rolling out networks in parallel. Potential gains from cooperation at the investment stage are therefore evident.

Internationally there is considerable variation with respect to whether mobile firms cooperate in this dimension. Sweden is a particularly interesting case.[12] In December 2000 four 3G licenses were issued in Sweden based on a beauty contest. All the firms being awarded a license promised very aggressive network investments.[13] The dominating firm, Telia, was not that aggressive and was accordingly not granted a license. Soon after the licenses were issued, Telia formed a joint venture with the second largest mobile firm, Tele2. The joint venture, called Svensk UMTS nät AB, is now rolling out a 3G network based on the license awarded to Tele2. Both Telia and Tele2 offer 3G services to end users based on capacity from the joint network. The Swedish Competition Authority (2002) approved the cooperation under a set of conditions. The approval is time limited up until February 2007. The Swedish case implies an extreme level of cooperation at the investment stage. Less extreme examples of cooperation are Germany and the UK. Taking Germany as an example, the European Commission[14] has approved the sharing of sites and also national roaming for a limited time period.

The Commission, in its decision on roaming in Germany, as well as the competition authorities in Sweden[15] try to balance gains from cooperation against the possible adverse effects on competition in the end-user market. In both cases the approval of cooperation is time limited. From the decisions on UK and Germany it is quite explicit that the firms will not be allowed to

---

[11] The standard for 3G mobile networks being deployed in Europe is called UMTS; Universal Mobile Telecommunications System, whereas GSM (Global System for Mobile Communications) is the 2G standard.

[12] See Hultén et al. (2001) for a description of the 3G license process in Sweden

[13] As an example, according to Hultén et al Europolitan and HI3G planned to build 700 base stations a month.

[14] See European Commission 2003

[15] See Swedish Competition Authority 2001

cooperate on roaming after the approval expires. This is in contrast to Sweden.

4. Modelling issues

In this section we will discuss some modelling issues of importance when analysing competition between networks. These issues are related to network externalities and their implications.

4.1. Externalities and the generalised price of communication

Consider a consumer with N potential communication partners (friends, family, colleagues etc.). Let q_i (i = 1,2,…N) denote the volume of communication with partner i and let p_i denote the quality adjusted generalized price of communication over the most efficient available network. The generalized price contains all relevant aspects of costs attached to communication, e.g. quality (oral, writing, face to face), cost of traveling (if communication is face to face), waiting (traditional mail), etc. Let u denote the net utility from communication; $u = \max\left[U(q_1,...q_N) - \sum p_i q_i\right]$. If consumer number i for some reason chooses to join a network with the effect that the generalized price goes down (up), then the net utility from communication will increase (decrease). This is a network externality since consumer i's choice of joining a network or not affects the utility of consumer j. Note that this externality is also present in more generalized frameworks where the consumer also optimizes over which available network is being used each time. An indirect utility function is always decreasing in price. Prices are determined by the decisions of joining networks by other consumers. The presence of network externalities as such does not therefore depend upon restrictive assumptions used when modelling.

Consider a simplified example where the intensity of communication is identical over all communication partners.[16] Assume there are two networks available, mail and telephony, all communication partners are members of the mail network and some are also members of the telephony network. The generalised price of telephony is assumed to be lower than the price of mail. Then, for given generalised prices, the utility of a consumer can be written as

[16] In this section we assume that if the generalized price is the same towards all communication partners, then the call quantity is also identical. In the literature this is called a uniform calling pattern, see Laffont Rey and Tirole (1998a).

$u(n)$, where n is the number of consumers that have joined the telephony network. This function is illustrated below:

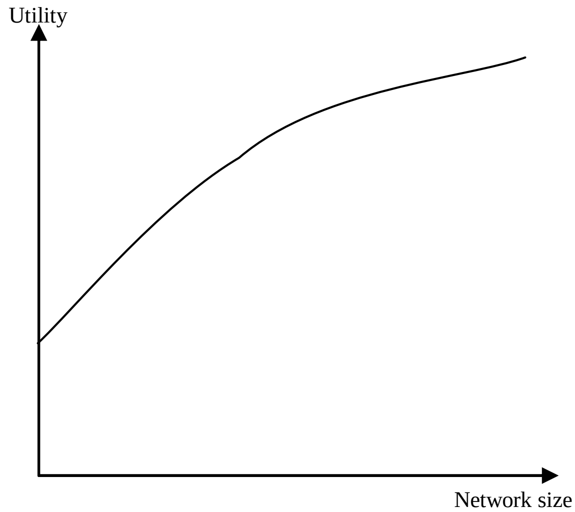

Figure 6, Utility as a function of network size

This type of utility function is applied in a number of papers studying network externalities, see e.g. Katz and Shapiro 1985, where it is assumed concave as illustrated above.

In models of network competition based on the Laffont Rey and Tirole (1998a,b) framework, the externality function is assumed linear. Thus when the price of on- and off-net communication is differentiated or participation is partial (à la Dessein 2003) utility varies linearly with the number of consumers on the network. Under these assumptions, the total value of the network is proportional to the square of the network size. Thus one assumes that "Metcalfe's law" is fulfilled.[17] This property is by many considered unlikely to hold because consumers are heterogeneous. In Rohlfs (2005) section 2.4 it is argued that consumers indeed are heterogeneous and that the ones that are most communication intensive will join first. Rohlfs argues that: "*the value of a network increases much less than proportionately to the square of the number of users*". A more realistic modelling of network effects where the utility function is concave in network size will however not necessarily qualitatively change the results from models on network competition. It is the incentives facing marginal consumers (the ones that are

[17] According to Metcalfe's law the value of a network increases proportionally to the square of the number of users. Bob Metcalfe is the inventor of Ethernet.

indifferent as to joining the network or not) that determine the size of a network. As a local approximation the utility function can be assumed linear. According to Rohlfs the slope of the utility function at the margin is below the average slope. Thus one has to be careful not to overstate network effects.

4.2. Externalities and network competition

Consider now a fairly standard model of network competition. There are two competing networks, there is full participation (consumers are on one or the other network), calling patterns are uniform, network effects are linear and the size of the market is normalized to unity. Let p_i and $\hat{p}_i$ denote the price of making on- and off-net calls respectively, for consumers on network *i* (*i* =1, 2). Then the utility of being on network *i* can be written:

$$\alpha\omega(p_i)+(1-\alpha)\omega(\hat{p}_i)$$

where $\alpha \in [0,1]$ is the size (= market share) of network *i* and $\omega()$ is an indirect utility function. Note that if the prices for on- and off-net calls are identical, then network externalities disappear. Consumers are indifferent as to the choice of network made by others. Consider now the case where $\hat{p}_i > p_i$. Then net utility as a function of size of network *i* can be illustrated as below:

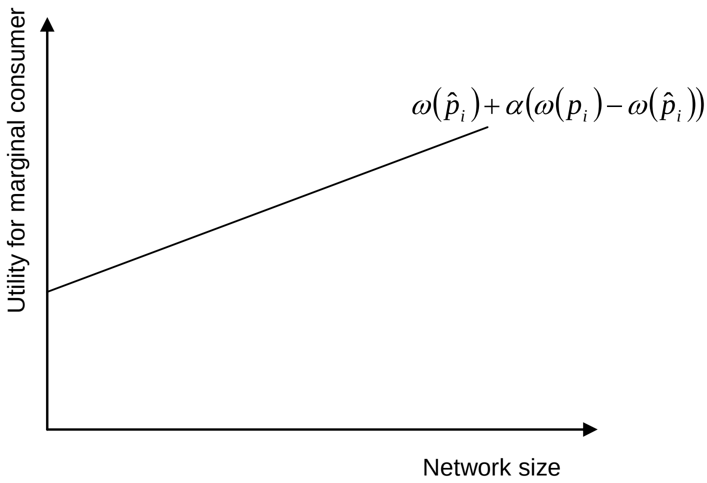

Figure 7, Linear network effects

In the Laffont Rey and Tirole (1998a,b) framework it is assumed that the two competing networks are differentiated à la Hotelling. Consumer preferences are assumed to be distributed uniformly on the unit interval, and the disutility from not consuming the most preferred variety is linear in distance from the location of preferences and the location of the chosen network. The net utility of joining network *i* located at x_i for a consumer of type *j*, with preferences located at x_j is accordingly:

(1.) $$\alpha\omega(p_i) + (1-\alpha)\omega(\hat{p}_i) - t\left|x_j - x_i\right| - T_i$$

where *t* is the linear disutility of not consuming the most preferred brand (travelling cost) and T_i is a fixed fee charged by network *i*. Consider the net utility of a consumer joining a network located at 0. It can be decomposed in three terms, a constant independent of network size, network effects and travelling costs: $\omega(\hat{p}_i) - T_i + \alpha(\omega(p_i) - \omega(\hat{p}_i)) - t\alpha$, the terms depending on network size is illustrated below:

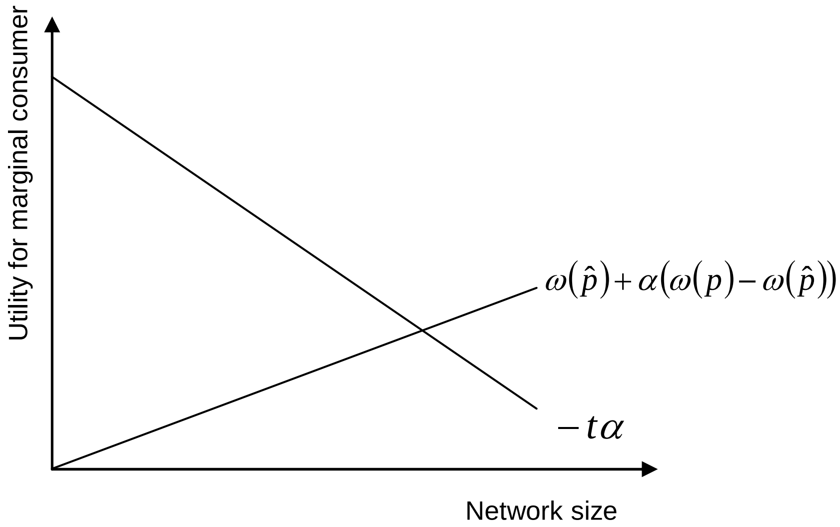

Figure 8, Net utility of the marginal consumer, decomposed

As is well known from Hotelling type models, profits are driven by the degree of horizontal differentiation. The larger the travelling cost *t* the steeper is the line -$t\alpha$ and the higher are profits. In this kind of models, network effects have the same implication as if the travelling cost is reduced. It is the sum of travelling cost and network effects that explains the net utility for the marginal consumer. Thus comparing two equilibria; one with large network effects and one with small network effects, profits will be higher in the latter. This mechanism drives the results in one of the papers in the present thesis, Foros and Hansen (2001), as well as some results in Farell and Saloner (1992) and Gans and King (2001). In section 5.3 of this introduction I will discuss our contribution in relation to the literature.

4.3. Externalities and the choice of strategic variable

Consider the net utility function (1.) above. We can define $V_i = \alpha\omega(p_i) + (1-\alpha)\omega(\hat{p}_i) - T_i$ as the surplus of joining network *i*. Then net utility can be written in two equivalent ways:

$$\alpha\omega(p_i) + (1-\alpha)\omega(\hat{p}_i) - t\left|x_j - x_i\right| - T_i$$
$$\equiv V_i - t\left|x_j - x_i\right|$$

On the one hand it seems reasonable to assume that firms compete by setting prices, on the other hand, it may be more convenient to consider net utility as the choice variable.[18] Since (indirect) utility is measured in money there seems to be a one-to-one relationship between price and utility. In Laffont Rey and Tirole (1998b) it is argued (p. 52): *Again we are back to a single-dimensional competition (competition in net surpluses or equivalently in fixed fees).”* Armstrong (2002) however argues that this claim is not necessarily valid. When solving a model similar to the LRT model he states, in a footnote on page 359, that: *“A subtle point is that one has to take care about the choice of strategic variables when network effects are present”*. Armstrong does not elaborate on this point however, and it may be worthwhile to take a closer look at this issue.[19]

Armstrong’s point can be illustrated by looking at a simplified[20] version of the stage 2 game in Foros and Hansen (2001). The simplifications are done to save notation and focus on the main aspect. Let the utility of being connected to network *i* be:

$$U_i = v + \alpha_i + k(1-\alpha_i) - t\left|x - x_i\right| - p_i$$

---

[18] See e.g. Armstrong and Vickers (2001)

[19] The issue came to my attention when writing the paper on asymmetric costs and network competition. Since there were two equivalent ways of solving the same problem my idea was to check my calculations by doing both. I was not able to obtain the same result.

[20] The simplification being that we do not take vertical differentiation into account, i.e. the parameter θ is set equal to zero. Furthermore the parameter β is set to unity (the degree of network externalities).

where v is the stand alone value of the network service. In this model network effects are linear. Total network size is normalized to unity, α_i is the market share of firm i, thus α_i is the value of being able to communicate with others on the same network. The parameter $k \in [0,1]$ measures the quality when consumers communicate with subscribers on the other network, thus $k(1-\alpha)$ is the utility from off-net communication.

Then net utility is defined as the utility of consuming the product, minus the price:

$$V_i = v + \alpha_i + k(1-\alpha_i) - p_i$$

The market share functions are derived by identifying the location of the indifferent consumer:

$$\alpha = \frac{1}{2} + \frac{1}{2t}(V_i - V_j),$$

or alternatively, if price is the strategic variable as:

$$\alpha_i = \frac{1}{2} - \frac{1}{2(t-(1-k))}(p_i - p_j).$$

We assume that production costs are normalised to zero. When price is the strategic variable, firms maximise: $\alpha_i p_i$. When net utility is the strategic variable we must substitute from the definition of net utility in order to eliminate price from the profit expression:

$$\alpha_i(v + \alpha_i + k(1-\alpha_i) - V_i).$$

4.3.1. Price as strategic variable

The first order condition for maximised profits is:

$$\frac{\partial \alpha_i}{\partial p_i^p} p_i^p + \alpha_i = 0 \Leftrightarrow -\frac{1}{2(t-(1-k))} p_i^p + \frac{1}{2} - \frac{1}{2(t-(1-k))}(p_i^p - p_j^p) = 0$$

where superscript p denotes that price is the strategic variable. In a symmetric equilibrium ($\alpha = 0.5$, $p^p = p_i^p = p_j^p$) prices and profits become:

$$p^p = t - (1-k)$$
$$\pi^p = \frac{t}{2} - \frac{(1-k)}{2}$$

4.3.2. Net utility as strategic variable

Maximising profits:

$$\max_{V_i}\left[\alpha_i\left(\alpha_i + k(1-\alpha_i) - V_i\right)\right] \Rightarrow 2\alpha_i(1-k)\frac{1}{2t} + (k - V_i)\frac{1}{2t} - \alpha_i = 0$$

In a symmetric equilibrium, net utility becomes: $V_i = 1 - t$. By inserting this expression back into the definition of net utility we find the equilibrium price and equilibrium profits:

$$p_i = \alpha_i + k(1-\alpha_i) - V_i \Leftrightarrow$$
$$p^V = t - \frac{1-k}{2}$$
$$\pi^V = \frac{t}{2} - \frac{1-k}{4}$$

where superscript *V* denotes that net utility is the strategic variable.

4.3.3. The two solutions compared

First of all, it is apparent from the calculations above that equilibrium prices and profits depend upon whether price or net utility is the strategic variable. In the model above this is the case if $k < 1$, i.e. if the model exhibits network effects. Prices and profits are higher when net utility is the strategic variable.

In order to investigate whether the difference is significant I have illustrated the best response functions numerically for the two solutions below:

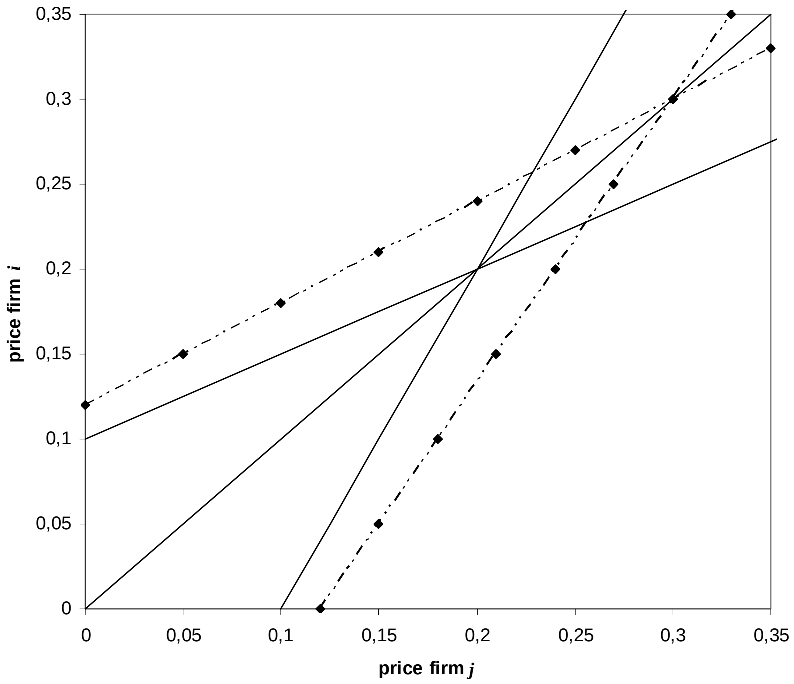

Figure 9, The best response functions compared

In this example the parameter values: t = 0.4, k = 0.8, have been used. The solid lines are best response functions when price is the strategic variable and the dotted lines are best response functions (in the pricing dimension) when net utility is the strategic variable. In the games we consider here the difference in outcome is significant, equilibrium prices are 0.2 and 0.3 respectively.

In the simple model we consider here, the parameter restrictions that ensure a shared equilibrium are: $t > (1 - k)$. Numerical simulations have revealed that parameter combinations close to this boundary[21] yields a large difference between the two equilibria relative to parameter restrictions well inside the boundary.

When firms have net utility as a strategic variable they commit to a less aggressive behaviour. This becomes apparent if we look at the locus of the

---

[21] One interpretation of the parameter restriction is that it assures that the network effect is not dominating the Hotelling differentiation. Thus the parameter restriction assures that the market share functions decrease in price.

best response functions at the previous page. For any price determined by the opponent, the best response price is higher in the game where net utility is the strategic variable as compared to the game where price is the strategic variable.

Consider a game where firms offer a contract where the level of the fixed fee is a function of market share, i.e. $p_i = A_i + b_i\alpha_i$ where, in the notation of the model above parameters are set such that: $A_i = v + k - V_i$ and $b_i = 1 - k$. Then it becomes apparent that there is not a one-to-one relationship between p_i and V_i since the price paid also depends on market share. Instead we can see that using net utility as a strategic variable is equivalent to using the parameter A_i as a strategic variable. Recall that network effects and horizontal differentiation have opposite effects on the willingness to pay for the marginal consumer as illustrated in figure 8 above. When net utility is used as a strategic variable the competing firms commit to neutralise the effect of market share on the willingness to pay by the marginal consumer. By doing so, profits increase.

In the present thesis, the choice of strategic variable has impact on the solution in both the paper on cost asymmetries (paper 2) and the paper on internet competition (paper 3). In both cases price is used as the strategic variable. According to Armstrong (2002), this is (perhaps) the most plausible assumption.

5. Summary of the dissertation

In this section the abstracts of the four essays are reproduced. Furthermore, where relevant, recent results from the literature are related to the papers.

5.1. Termination rates and fixed mobile substitution

In this paper we consider fixed mobile substitution in a model of mobile network competition. We demonstrate that the termination rates are profit neutral if the size of the mobile sector is given. An implication of this result is that the mobile termination rate does not have an impact on profits in the mobile sector if all subscribers multihome. Furthermore, the termination rate is also profit neutral if there is fixed mobile substitution of a type where consumers change status from multihoming in fixed and mobile to a status where they singlehome in mobile. In situations where consumers multihome and there is a positive termination margin, mobile firms will set usage prices above perceived marginal cost.

Furthermore, if fixed mobile substitution results in an increased number of mobile subscribers, then the mobile termination rate will have an impact on profits in the mobile sector. The mechanism behind this result is that profits in the mobile sector are proportional to the size of the mobile sector. The size of the mobile sector is an increasing function of the net utility offered to mobile subscribers. This net utility is increasing in the termination rate because termination revenues are being passed on to consumers due to competition in the mobile sector. Thus the mobile termination rate will have an impact on profitability in the mobile sector if the size of the mobile sector is affected.

In a mixed market situation where the size of the mobile sector is elastic and there are some subscribers multihoming, the two effects described above, will in combination result in two kinds of market distortions. In order to induce consumers to joint the mobile networks, mobile firms will set termination rates above cost. Then, given a margin on termination and the existence of multihoming subscribers, mobile firms will have an incentive to raise usage prices above perceived marginal cost in order to make multihomers substitute traffic originated in mobile for traffic originated in fixed because it results in increased termination revenues.

The implication of the analysis in this paper is that there is a strong case for regulating mobile termination rates in the growth phases of mobile telephony, whereas there is less need for regulation in mature markets characterized by a stable size of the mobile sector. This seems to be the opposite of the approach taken by regulators in Europe, where mobile firms were free to set termination rates in the growth phase and where regulation is introduced once markets mature. The observed policy may however be explained by regulators wanting to stimulate the growth of the mobile sector. This policy is evidently resulting in reduced welfare (in the short run).

5.2. Network Competition when Costs are Heterogeneous

In this paper we study network competition when costs differ among two interconnected networks. We analyze the implications of three different principles for regulating termination fees when marginal costs differ. The first case we analyze is cost based in the sense that termination fees exactly reflect marginal costs. It is a standard result in the literature that usage prices then are determined at the optimal level. We demonstrate that with cost differences, equilibrium market shares are not optimal in this regime. The most efficient network is too small compared to a welfare maximizing solution. The reason is that with cost differences there is a tariff mediated network externality. There is however no mechanism in the market that enables the efficient firm to internalise this effect.

In the second regulatory regime we consider taxation and subsidisation respectively, of the two firms based on the number of subscribers as an addition to the cost based regulation of termination rates. By subsidising the low cost firm and/or imposing a tax on the high cost firm, the regulator can implement first best.

In the third regime we investigate whether granting a termination mark-up to the low cost firm can improve the situation as compared to cost based regulation. We demonstrate that the mark-up has the desired effect on market shares; the low cost firm becomes bigger. Furthermore, we demonstrate that, starting from cost based regulation, welfare increases as a termination mark-up granted to the low cost firm is introduced. Thus it is welfare improving to let the efficient firm enjoy a (small) mark-up.

.The results described above are derived within a model not taking into account that consumers may derive utility from receiving calls. If the opposite is the case, then consumers, in their choice of network, also will take into account how much it will cost for others to call them. Thus, taking receiver utility into account, it may result in a reduced welfare loss due to the low cost firm being to small under cost based regulation. Asymmetric models with receiver utility is not possible to solve analytically according to Hoernig (2006).

Jeon, Laffont and Tirole (2004) analyse a model of network competition with symmetric costs where consumers also derive utility from receiving calls. They demonstrate that if the utility of receiving calls is identical to the utility from originating calls, then networks will set infinitely high off net prices. The result is connectivity breakdown. The mechanism driving this result is that the competing firms attempt to offer a bundle superior to their competitor. When the utility of receiving calls is significant, the utility of customers in the competing network increase fast in the volume of off-net traffic. Thus by increasing the off net price customers on the other network is "punished". The pricing structures observed in the telephony market with relatively moderate on-net off-net differentiation is accordingly not consistent with a large utility of receiving calls.

5.3. Competition and compatibility among Internet Service providers

Information Economics and Policy 13 (2001) 411–425

Co-authored with Øystein Foros, Norwegian School of Economics and Business Administration

We consider a two-stage game between two competing Internet Service Providers (ISPs). The firms offer access to the Internet. Access is assumed to be vertically and horizontally differentiated. Our model exhibits network externalities. In the first stage the two ISPs choose the level of compatibility (i.e. quality of a direct interconnect link between the two networks). In the second stage the two ISPs compete à la Hotelling. We find that the ISPs can reduce the stage 2 competitive pressure by increasing compatibility due to the network externality. The firms will thus agree upon a high compatibility at stage 1. When it is costly to invest in compatibility, we find that the firms overinvest, as compared to the welfare maximising investment level.

Competing firms will accordingly have incentives to reduce network effects by decreasing the on – off net quality differential. Similar insight was to my knowledge first developed by Farrell and Saloner (1992). Farrell and Saloner analyse competition between two technologies. Both technologies are characterized by network effects, and if the technologies are compatible or there are converters available, network effects will also flow across the networks. They assume full participation, thus if converters are costless and perfect, network effects disappear similar to the case in our paper when there is no on-net, off-net quality differentiation. Under duopoly Farrell and Saloner find that profits increase in the quality of the converter. In the absence of converters, the equilibrium derived by Farrell and Saloner is symmetric. This is in contrast to the conversion equilibrium which is asymmetric. Due to a mechanism outside the Farrell Saloner model, consumers expect one of the two technologies to become dominant. Consumers buying the dominant technology do not buy a converter, whereas consumers buying the dominated technology also buy a converter.[22] With respect to converters, both firms want converters to have a high quality. The

---

[22] In the Farrell Saloner model, a conversion equilibrium cannot be symmetric. Consider a consumer with strong preferences for one of the technologies. If consumers on the other technology buy converters, the best response is not to buy a converter, and vice versa. In equilibrium the market share of the dominant technology exceeds the market share of the dominated technology.

dominant firm wants converters to be expensive, whereas the dominated firm wants them to be cheap. There are some notable differences between our model and the Farrell Saloner model. In contrast to Farrell and Saloner, we consider vertical differentiation, suppliers share the cost of compatibility, and this functionality is bundled into the product.

Roson (2002) dedicates an entire paper to comparing our paper on interconnection on the internet to a paper by Cremer et al. (2000) on the same issue. Both papers discuss the incentives to interconnect on the internet but arrive at opposite conclusions. Cremer et al. find that dominant firms may have incentives to degrade interconnection. There are a number of differences with regard to assumptions between the two papers. Roson argues that the difference in conclusions is primarily driven by different assumptions regarding market size. We assume that the size of the total market is given whereas Cremer et al. assumes that market size is a function of prices. As demonstrated in another paper in this thesis (the paper on fixed mobile substitution) the incentives to interconnect are indeed dependent upon whether total market size is given or not. Cremer et al. however also assume that there are two groups of consumers; captured consumers, not responding to price changes, and a group of noncaptured consumers responding to price changes. Economides (2005) analyzes the implication of the assumed captured customers. He demonstrates that if all consumers respond to price changes, i.e. that the number of captured consumers is zero, then the conclusions of Cremer et al. change, and they become in line with the result in our paper: Networks have a common interest in assuring a high interconnection quality.

5.4. Demand-side Spillovers and Semi-collusion in the Mobile Communications Market

Journal of Industry Competition and Trade, 2002, 2, (3), pp. 259-278

Co authored with Øystein Foros, Norwegian School of Economics and Business Administration and Jan Yngve Sand, University of Tromsø

We analyze roaming policy in the market for mobile telecommunications. Firms undertake quality improving investments in network infrastructure in order to increase geographical coverage, capacity in a given area, or functionality. Prior to investments, roaming policy is determined. We show that under collusion at the investment stage, firms' and a benevolent welfare maximizing regulator's interests coincide, and no regulatory intervention is needed. When investments are undertaken non-cooperatively, firms' and the regulator's interests do not coincide. Contrary to what seems to be the regulator's concern, firms would decide on a higher roaming quality than the regulator. The effects of allowing a virtual operator to enter are also examined. Furthermore, we discuss some implications for competition policy with regard to network infrastructure investment.

In their review article on Wireless communications, Gans et al. (2005) base their discussion of the implications of national roaming on results from our paper.

The quality improvement stemming from investments in mobile networks can take the form of improved capacity and/or improved coverage. Our paper focuses on capacity. This is in contrast to Valletti (2003) where the emphasis is on coverage as a means to vertical differentiation. The duopoly equilibrium in the Valletti model is characterised by maximum differentiation. One firm chooses maximum coverage, the other chooses minimum coverage (minimum coverage is typically specified in the license). In the Valletti model, national roaming is unprofitable for the firms. Thus roaming is only profitable if the firms collude. This result is in contrast to our paper. In their review article Gans et al. (2005) argue that the Valletti result is due to simplifying assumptions.[23] Furthermore, observed market behaviour indicates that mobile firms tend to set similar coverage.

The market experience reviewed in chapter 3.4 in this introduction revealed that mobile firms in several countries indeed cooperate over roaming and investments. Furthermore, the regulating authorities, given a set of conditions, have approved the cooperation. Given the approach taken by the regulating authorities an interesting issue to analyse would be the implications of allowing cooperation only in a limited time period.

6. Concluding remarks

The tension between cooperation and competition is not a transitory phenomenon during a deregulatory process. Even under full facility-based competition there is a potential gain from interconnecting networks. In section 3 of this introduction I have argued that interconnection is an issue for “old” as well as “new” services. The four papers in the present thesis indicate that the costs and benefits of interconnection and thus the incentives to interconnect change with technology and market characteristics. There is accordingly a need for a case by case analysis when assessing the need for regulatory intervention in such markets.

[23] Gans et al. (2005, page 270) argue that the Valletti result is due to the assumed pure vertical differentiation. If there is horizontal differentiation in addition to the vertical differentiation, then firms may set identical coverage and instead compete in other dimensions.

7. Literature

Armstrong, M, 2002, The theory of access pricing and interconnection, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 1, pp 295-384 (Elsevier science, Amsterdam).

Armstrong, M, J. Vickers 2001, Competitive price discrimination, Rand Journal of Economics, Vol 32 no 4, pp 579 – 605.

Armstrong, M., 1998, Network interconnection in telecommunications, The Economic journal, 108, (May), 545 – 564.

Brock, GW. 2002, historical overview, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 1, pp 43-74 (Elsevier science, Amsterdam).

Carter, M. and J. Wright, 2003, Asymmetric Network Interconnection, Review of Industrial Organization, 22, 27 – 46.

Cremer, J., P. Rey, and J. Tirole, 2000, Connectivity in the Commercial Internet, The Journal of Industrial Economics, Vol. 48, No. 4, pp. 433-472.

D'Aspremont, C. and A. Jacquemin, 1988, Cooperative and Noncooperative R&D in Duopoly with Spillovers, The American Economic Review, Vol 78, no 5, pp 1133 – 1137.

DeBondt, R. 1997, Spillovers and innovative activities, International Journal of Industrial Organization, Vol. 15, no 1, pp 1 – 28.

DeGraba, P. 2003, Effcient Intercarrier Compensation for Competing Networks When Customers Share the Value of a Call, Journal of Economics & Management Strategy, Vol. 12, No. 2, pp 207–230.

Dessein, W. 2003, Network Competition in Nonlinear pricing, Rand journal of economics, vol 34, no 4, pp 539 – 611.

Economides, N. 2005, The Economics of the internet Backbone, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 2, pp 373-412 (Elsevier science, Amsterdam).

European Commission, 2003, Commission approves 3rd Generation mobile network sharing in Germany, IP/03/1026, Date: 16/07/2003, available at: http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/03/1026&format=HTML&aged=0&language=EN&guiLanguage=en.

Farrell, J. and G. Saloner, 1992, Converters, Compatibility, and the control of Interfaces, Journal of Industrial Economics, vol. 40, no. 1, pp. 9-36.

Faulhaber , GR. 2003, Policy-induced competition: the telecommunications experiments, Information Economics and Policy, Vol 15, 73–97.

FCC, 1998, Memorandum Opinion And Order, In the Matter of Application of WorldCom, Inc. and MCI Communications Corporation for Transfer of Control of MCI Communications Corporation to WorldCom, Inc., FCC 98-225, Washington DC

Gans, JS. and SP. King, 2001, Using ’bill and keep’ to interconnect arrangements to soften network competition, Economic Letters, 71, 413 – 420.

Gans, JS. et. al, 2005, Wireless communications, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 2, pp 243-285 (Elsevier science, Amsterdam).

Hoernig, S. 2006, On-Net and Off-Net Pricing on Asymmetric Telecommunications Networks, CEPR discussion paper 5588

Hultén, S. et al. 2001, 3g Mobile Policy: The Case Of Sweden, ITU, available at: http://www.itu.int/osg/spu/ni/3G/casestudies/sweden/Sweden_fin3.pdf

Jeon, DS, J-J. Laffont and J. Tirole 2004, On the “receiver pays” principle, Rand Journal of economics, vol 35, no 1, pp 85 110

Katz, ML and C. Shapiro, 1985, Network externalities, Competition and Compatibility, The American Economic Review, Vol 75, no 3, pp 424 - 440

Kende, M. 2002, The Digital Handshake: Connecting Internet Backbones, Federal Communications Commission, Washington DC, OPP Working Paper No. 32

Laffont, J-J and J. Tirole, 2000, Competition in telecommunications, (The MIT press, Cambridge, Mass)

Laffont, J-J., P. Rey, J. Tirole, 1998a, Network Competition: I. Overview and non-discriminatory pricing, RAND Journal of Economics, 29, (1) 1 – 37

Laffont, J-J., P. Rey, J. Tirole, 1998b, Network Competition: II. Price discrimination, RAND Journal of Economics, 29, (1) 38 – 58

Liebowitz, SJ. and SE. Margolis, 2002, Network Effects, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 1, pp 75 - 96 (Elsevier science, Amsterdam)

Matutes, C. and P. Regibeau, 1988, Mix and Match: Product Compatibility

Without Network Externalities, Rand Journal of Economics, vol. 19 (2), pp. 219-234

NPT 2005a, Varsel om vedtak om utpeking av tilbydere med sterk markedsstilling og pålegg om særskilte forpliktelser i markedene for originering, terminering og transitt av offentlig telefontjeneste i fastnett, 15. september 2005, available at http://www.npt.no/iKnowBase/Content/varsel_8_10_15092005.pdf?documentID=45500

Peitz, M. 2005, Asymmetric access price regulation in telecommunications markets, European Economic Review, 49, 341 – 358.

Rinde, H, 2005, Et telesystem tar form 1885 – 1920, Norsk telekommunikasjonshistorie bind I, (Gyldendal, Oslo)

Rochet, JP., J. Tirole, 2003, Platform Competition in Two-Sided Markets, Journal of the European Economic Association, vol. 1, n. 4, June 2003, p. 990-1029.

Rohlfs, JH. 1974, A theory of interdependent demand for a communications service, Bell Journal of Economics and Management Science, Vol 5, no 1, pp 16- 37

Rohlfs, JH. 2005, Bandwagon effects in Telecommunications, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 2, pp 81-118 (Elsevier science, Amsterdam)

Roson, R. 2002, Two Papers on Internet Connectivity and Quality, Review of Network Economics, Vol. 1, no 1, pp. 75 - 80

Shapiro, C., H. Varian, 1998, Information Rules: A Strategic Guide to the Network Economy, Harvard, Business School Press, Boston, Massachusetts.

Swedish Competition Authority, 2002, Undantag - samarbete om UMTS-nät, Beslut 2002-03-12, Dnr 321/2001, available at: http://www.kkv.se/beslut/01-0321.htm

Valletti, T., 2003, Is Mobile Telephony a Natural Oligopoly?, Review of Industrial Organization no 22, pp 47–65

Section 2

Termination rates and fixed mobile substitution

Termination rates and fixed mobile substitution*

By:
Bjørn Hansen
Department of Economics
BI, Norwegian School of Management
N-0442 Oslo
Norway
Email: bjorn.hansen@bi.no

April 2006

Abstract:
In this paper we consider fixed mobile substitution in a model of mobile network competition. We demonstrate that the profit neutrality result from the standard model of network competition (Laffont Rey Tirole1998a) holds if the number of mobile subscribers is given. Thus the mobile termination rate does not have an impact on profits in the mobile sector in a mature market where all consumers are hooked up to a mobile network and fixed mobile substitution results in subscribers disconnecting from the fixed network. However, if fixed mobile substitution results in an increased number of subscribers in mobile networks, then the mobile termination rate will have an impact on profits in the mobile sector. The implication of the analysis is that there is a case for regulating mobile termination rates in the growth phase, whereas there is less need for regulation in mature markets characterized by a stable size of the mobile sector. This seems to be the opposite of the approach taken by regulators in Europe, where mobile firms were free to set termination rates in the growth phase and regulation of termination rates is introduced once markets mature.

* *I would like to thank Christian Riis and Øystein Foros for valuable comments to earlier versions of this paper.*

1. Introduction

The demand for telephony services is derived from the need for communication. This communication will take place either from a fixed or a mobile phone. Intuitively one would accordingly expect fixed and mobile telephony to be substitutes. The implications of fixed mobile substitution on the regulation of mobile termination rates should thus be taken into account. To do this we extend the model of network competition due to Laffont Rey and Tirole (1998a) by adding a fixed network. We show that a necessary condition for the profit neutrality of mobile termination rates is that the size of the mobile sector is given. If, on the other hand, the size of the mobile sector is elastic, i.e. the growth phase of mobile telephony, then the mobile firms can increase their profits by raising the termination rate.

There is empirical evidence supporting the notion that fixed and mobile services are substitutes, see Cadima and Barros (2000) and Gruber and Verboven (2001b). There are however also examples of studies finding that fixed and mobile services are complements, e.g. Gruber and Verboven (2001a). According to a review article by Gans King and Wright (2005), the measured complementarity may be explained by network effects in the early phases with relatively few mobile subscribers. They argue that fixed and mobile services are likely to be substitutes in the mature phases of the life cycle of mobile telephony.

In order to obtain telephony connectivity, networks have to be interconnected, i.e. two-way access is required. Under the widespread principle of calling party pays, the interconnecting networks both buy and sell termination services. The termination service is accordingly an input to other phone companies, both fixed and mobile. The literature on interconnection of symmetric networks, i.e. mobile to mobile termination, starting with the papers by Laffont Rey and Tirole (1998a) as well as the paper by Armstrong (1998) is inconclusive with respect to whether interconnecting, competing firms have incentives to set termination rates above the welfare maximizing level. The results depend upon the pricing structure in the downstream market. The result that interconnected networks do not necessarily have incentives to set termination rates above the welfare maximizing level is in contrast to the results in the literature on fixed to mobile termination. Fixed companies are typically former monopolies and they are subject to regulation. The argument is that fixed companies do not have any bargaining power when negotiating termination rates due to regulation. Thus mobile firms have incentives to raise termination rates above the welfare maximizing level in order to extract profits from the fixed sector (see e.g. Armstrong 2002 p. 339 and Wright 2002). A review of results on fixed to mobile termination can be found in de Bijl et al. (2004). None of these papers consider the implications of fixed mobile substitution.

The results cited above for mobile to mobile termination and fixed to mobile termination respectively are not derived within the same modelling frameworks. From the outset it is accordingly not evident whether the differences in conclusions with respect to the incentives to set termination rates survive within a model studying both issues simultaneously.

We consider a model where competing mobile firms set a termination rate that also applies to the fixed network. In the paper we demonstrate that the mobile firms pass termination revenues on to their subscribers due to competition in the mobile sector. On the one hand, the passing on of revenues implies that the profit neutrality result from the network competition literature (first derived in Laffont Rey and Tirole 1998a) still is valid given that the number of subscribers on mobile networks is given. On the other hand, the passing on of revenues implies that if the demand for mobile subscriptions is elastic, i.e. consumers substitute mobile for fixed, then mobile firms have incentives to raise their termination rate above the welfare maximizing level in order to attract more subscribers. In a mobile market in growth ("emerging market") mobile firms can accordingly increase profits by raising the termination rate, whereas in a saturated mobile market ("mature market") the profit neutrality of termination rates holds. To our knowledge, our analysis of fixed mobile substitution as well as multihoming is novel.

In the present paper we allow for multihoming, we analyze the effect of fixed and mobile services being substitutes in consumption and we assume non discriminatory termination rates, i.e. that mobile firms charge the same termination fee regardless of whether the traffic is originated in a fixed network or a mobile network. To fix ideas we can think of two competing mobile firms negotiating a reciprocal termination fee,[1] and then due to the non discrimination assumption apply the same termination rate towards the fixed network. Interconnection is illustrated below:

[1] *We assume that the reciprocal mobile termination fees are determined in negotiations between mobile firms, or by the regulator. The outcome of (symmetric) mobile firms setting termination rates non-cooperatively is also reciprocity as demonstrated by Gans and King (2001). However, when termination rates are determined non-cooperatively, Gans and King demonstrate that equilibrium level is high relative to the outcome under cooperative determination of termination rates. In many jurisdictions, regulators set the mobile termination rates. Regulated rates are reciprocal in some countries and non reciprocal in others.*

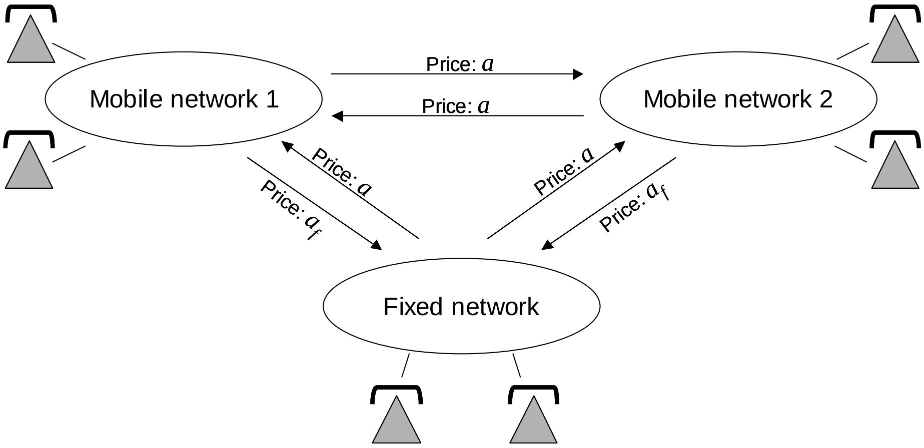

Figure 1, Termination and non discrimination

In the illustration above the reciprocal mobile to mobile termination fee is denoted a. The fixed network has to pay the same termination fee to both mobile networks. Furthermore, a_f denotes the termination fee charged by the fixed network. In European markets the mobile to mobile termination fees are typically 5 to 10 times higher than the termination fee charged by fixed networks. This is contrary to the US where reciprocal fixed to mobile termination rates are observed.

The assumed non discrimination is common among mobile operators (see de Bijl et al. 2004 p 108). This assumption is critical for our results and can be motivated in two ways; due to regulation and/or arbitrage. In many jurisdictions (e.g. most of the EU) non discrimination is mandatory on the termination market; i.e. mobile firms are not allowed to price discriminate based on whether the calls are originated in a fixed or a mobile network. Furthermore, suppose discrimination is allowed, then calls from the network facing the high termination charge can be routed via a network facing a low termination fee and thus the price discrimination is bypassed.[2]

The current paper is organized as follows. In section 2 of the paper we reproduce the reference model, i.e. a model of network competition with two part tariffs in the downstream market. Within the reference model we demonstrate that the profit neutrality result also holds if we add a fixed

---

[2] *In the industry this type of bypass is called refiling. In Norway we had a case of refilling in 1999 – 2000, because the mobile firms had differentiated "domestic" and "international" termination fees. The international termination fees were below the domestic and calls were routed via Sweden in order to be subject to the lower international termination fee. Due to this bypass the differentiation of termination fees was abandoned.*

network of exogenous size. In section 3 of the paper we present our model of network competition and fixed mobile substitution. Then we proceed in section 4 by analyzing two types of market equilibriums, full multihoming and full singlehoming respectively. Finally in section 6 we conclude the paper. In appendix B we characterize two more possible equilibriums.

2. Adding an exogenous fixed network to the standard model of network competition

Laffont Rey and Tirole (1998a) presented a model with Hotelling type differentiation between two mobile firms where the networks charge two part tariffs. A striking result from this model is the profit neutrality of reciprocal termination charges. In this section of the paper we will add a fixed network of exogenous size to a model of the Laffont Rey Tirole type. This model will serve as a benchmark and a motivation for the models where we take fixed mobile substitution into account.

The mobile market is assumed to have a given size normalized to unity. Subscribers are assumed to single home, calling patterns are assumed to be uniform and the two competing mobile networks charge two part tariffs. The networks are differentiated à la Hotelling. Network preferences are assumed to be uniformly distributed on the unit line, and the differentiation is assumed exogenous. The utility for a subscriber of type *x* connected to network *i* is given by:

$$V_i + v_0 - \frac{1}{2\sigma}\left|x - x_i\right|$$

Where $V_i = \omega(p_i) - T_i$ and $\omega(p_i) = \max[u(q_i) - pq_i]$, is the net utility from network subscription, *q* is the number of calls being made, $u(q)$ is the utility form making calls and T_i is the fixed part of the two part tariff (subscription fee). Furthermore, v_0 is the stand alone value of subscription to a mobile network, x_1 (=0,1) is the locus of the two networks. The disutility from not consuming an offering of the preferred type is $1/2\sigma$. Market shares, α_i, are determined by the subscriber being indifferent as to the two offerings, thus:

$$\alpha_i = \frac{1}{2} + \sigma\left(V_i - V_j\right)$$

In this section we simplify the modelling by assuming that fixed subscribers make calls to the mobile subscribers, but mobile subscribers do not make calls to the fixed network.[3]

We assume that the volume of incoming calls per mobile subscriber from the fixed network is a decreasing (non increasing) function of the termination rate *a*: $q_0 = q_0(a)$. Given a market share of α_i, mobile firm *i* will then receive incoming F2M traffic: $\alpha_i q_0(a)$. Then we can write profit of mobile firm *i*:

$$\pi_i = \alpha_i\left(T_i + q_0(a)(a - c_0) + q(p_i)\left(p_i - c - \alpha_j(a - c_0)\right)\right) + \alpha_i\alpha_j q(p_j)(a - c_0)$$
.

Where *c* is the unit price of producing a call, and c_0 is the cost of terminating a call.[4] We use net utility as the strategic variable. Thus we substitute: $T_i = \omega(p_i) - V_i$. Furthermore we define $R(a) \equiv q_0(a)(a - c_0)$ as the per subscriber net revenue from incoming calls from the fixed network. Then the profit function can be written:

$$\pi_i = \alpha_i\left(\omega(p_i) - V_i + R(a) + q(p_i)\left(p_i - c - \alpha_j(a - c_0)\right)\right) + \alpha_i\alpha_j q(p_j)(a - c_0)$$

This function is to be maximized with respect to marginal price *p* and net utility *V*. For given net utilities market shares are given and we obtain the by now familiar result:

(1.) $$\frac{\partial \pi_i}{\partial p_i} = 0 \Leftrightarrow p_i = c + \alpha_j(a - c_0)$$

Net profit from carrying traffic is accordingly zero. Consider next: $\frac{\partial \pi_i}{\partial V_i} = 0$, substituting for usage pricing at marginal cost and then solving with respect to V_i yields:

---

[3] *This assumption simplifies the modelling without changing the main insights, later on in the paper we will consider a richer model.*

[4] *In contrast to the Laffont Rey Tirole model, we do not include fixed cost per subscriber nor volume dependent costs in the transmission network. These parameters are not the focus of the current paper.*

$$V_i = -\frac{\alpha_i}{\sigma} + \omega(p_i) + R(a) + \alpha_i q(p_i)(a - c_0) + q(p_j)(1 - 2\alpha_i)(a - c_0)$$

In a symmetric equilibrium ($\alpha_i = 0.5, \quad p_i = p_j$), the expression above simplifies to:

(2.) $$V_i = -\frac{1}{2\sigma} + \omega(p_i) + R(a) + \frac{1}{2} q(p_i)(a - c_0)$$

Optimal pricing and optimal net utility can be substituted back into the profit function and then we obtain:

$$\pi_i = \frac{1}{4\sigma}$$

Hence profits are unaffected by the termination charge. Under Hotelling competition, any revenue (loss) from termination is passed on to consumers (see 2.).

Prior to the market game analyzed above, the mobile firms may negotiate a reciprocal termination charge. Since profits are unaffected, mobile firms are indifferent with respect to the level of this termination charge and they will (weakly) prefer to set it at the welfare maximizing level. A positive margin on termination to mobile networks will thus result in transfers from fixed to mobile subscribers.

Wright (2002) as well as Armstrong (2002) analyze the effects of mobile termination rates in a model with an exogenous fixed network of the type considered in above. In contrast to our result, both conclude that the mobile sector can increase profits by raising the termination rate. Wright (2002) is however also deriving a similar result to ours, but he argues that it is a special case. He focuses on cases where mobile firms set termination rates individually towards the fixed sector and/or situations with a less competitive mobile sector. Armstrong (2002), (p. 337 and onwards) also analyzes the implications of mobile firms setting fixed to mobile termination rates individually, but he assumes perfect competition in the mobile sector.

3. A model of network competition and fixed mobile substitution

In this section we will describe the extensions made to the model in order to analyze the effects of fixed mobile substitution.

3.1. Preferences in the fixed mobile dimension

As argued in the introduction to this paper, there is reason to believe that fixed and mobile services are substitutes.

Furthermore, in the market one can observe some consumers singlehoming in mobile, others singlehoming in fixed and some consumers multihoming in the sense that they subscribe to both fixed and mobile services. Taking Norway as an example, the number of mobile subscriptions exceeds the number of inhabitants[5], and 83% of all households are hooked up to the fixed network. Most people in Norway are accordingly multihoming.[6] Furthermore, there seems to be a trend that consumers disconnect from the fixed network and become singlehomers in mobile. This phenomenon is called fixed mobile substitution.[7] Some predicts that this development will accelerate.

Our modelling of preferences in the fixed mobile dimension takes as its starting point that consumers differ in the degree that they are on the move. Some consumes are at fixed locations almost all the time and thus close to a fixed phone. Such consumers are assumed to have relatively low willingness to pay for mobile services. This is in contrast to people being mostly on the move. Such consumers have to rely on the mobile phone to be able to communicate, thus they have relatively high willingness to pay for being connected to a mobile service. Since fixed services typically are considerably cheaper than mobile services some consumers may even find that they are best off by multihoming, i.e. by placing calls in the mobile network only when they are away from a fixed phone.

In our model the total number of customers is normalized to unity. We let every consumer be characterized by two parameters, (x, y) uniformly and independently distributed on the unit square. x measures preferences in the mobile dimension, i.e. the locus of preferences on the Hotelling line in the same way as in the model reviewed above. y is a measure of preferences in

---

[5] According to the Norwegian Post and Telecommunications Authority there were 104 mobile subscribers per 100 inhabitants in Norway in 2005.
[6] June 2005, source: NPT 2005
[7] See also ITU 2003

the fixed mobile dimension. This parameter can be given a straightforward interpretation, a consumer of type (x, y) is on the move and thus away from a fixed phone a fraction of time equal to y. The unit square is illustrated below:

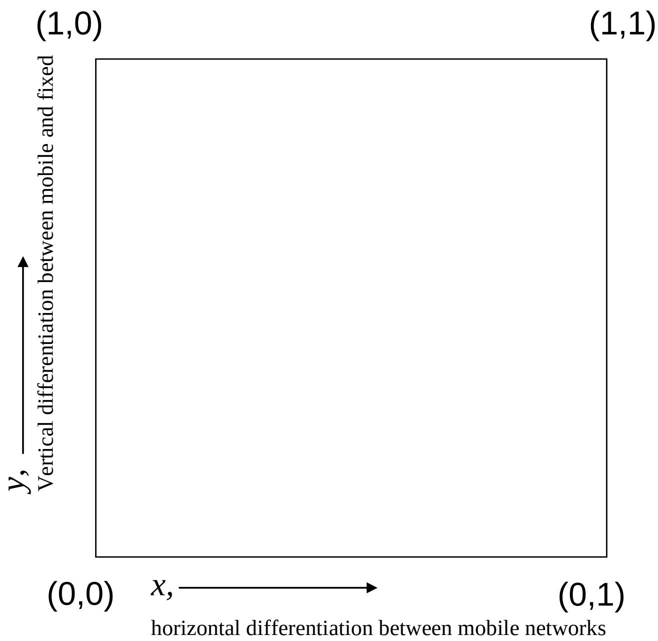

Figure 2, Product differentiation

Consumers with taste parameters in the upper left corner are likely to connect to mobile network 1, consumers with taste parameters in the lower half of the square are likely to connect to the fixed network, etc.

Note that differentiation between the two mobile services is assumed to be purely horizontal whereas the differentiation in the fixed to mobile dimension is purely vertical. Vertically differentiated mobile networks were analyzed by Carter and Wright 2003, as well as by Peitz 2005. As for the fixed mobile dimension, it seems reasonable to assume that fixed and mobile services are vertically differentiated since a mobile phone gives the opportunity of communication in fixed locations as well as the opportunity to communicate while being on the move. Some may however argue that there is an element of horizontal differentiation since mobile services are characterized by radiation, poorer sound quality and hassle related to charging batteries. Thus, alternatively one could model horizontal differentiation in the fixed to mobile dimension as well. Altering the modelling in the present paper by assuming Hotelling type horizontal differentiation in the fixed to mobile dimension yields qualitatively identical results.

A consumer of type (*x*, *y*) single homing on mobile network *i* is assumed to receive utility:

(3.) $$\omega(p_i) - \frac{1}{2\sigma}|x - x_i| + g(y) - T_i$$

The only difference from the utility function we considered in section 2 of the paper is that we have substituted the fixed term v_0 for a type dependent term, $g(y)$ capturing the preferences for mobile services.[8] We assume that $g'(y) > 0$.

The fixed network is assumed to be regulated in both the up- and downstream market, and the usage price on fixed is assumed to be an increasing function of the termination rate that the fixed network has to pay to mobile networks. The fixed network charges a single two-part tariff without discriminating between fixed to fixed and fixed to mobile traffic. Thus the indirect utility of a subscriber singlehoming in the fixed network (notation related to the fixed network has subscript *f* throughout the paper) can be written:

(4.) $$V_f(a) = \omega_f(a) - T_f, \quad V_f' \leq 0$$

Finally, multihoming subscribers will place calls from the fixed network as well as calls from one of the mobile networks. These calls are terminated in fixed and mobile networks proportionally to the respective market shares in the same way as assumed in the reference model considered in section 2. A multihoming subscriber is assumed to derive gross utility from making calls; $U(\hat{q}, \hat{q}_f)$, where $\hat{q}$ is the quantity of calls originated in the mobile network and $\hat{q}_f$ is the quantity of calls originated in the fixed network.[9] The multihoming consumer will optimize call consumption resulting in an indirect utility function; $\hat{\omega}(p,a) = \max_{\hat{q}_i, q_f}\left[U(\hat{q}, \hat{q}_f) - p\hat{q} - p_f\hat{q}_f\right]$. Thus the

---

[8] *Recall that the parameter y is measuring the fraction of time being away from a fixed phone. Thus, instead of adding a mobility premium to a mobile service one, an equivalent approach is to instead add a cost depending on y to the utility from fixed subscriptions.*

[9] *Throughout the paper, variables and functions with a hat are related to multihoming consumers, i.e.* $\hat{q}$ *is the quantity of mobile to mobile calls for multihoming subscribers and* q *is the quantity of mobile to mobile calls made by singlehoming consumers.*

utility of a subscriber of type (*x, y*) connected to mobile network *i* and to the fixed network is given by:

$$(5.)\quad \hat{\omega}(p_i, a) - \frac{1}{2\sigma}|x - x_i| + \hat{g}(y) - T_i - T_f$$

i.e. the sum of the following terms: indirect utility from making calls, the disutility from not consuming the most preferred mobile brand, the type dependent utility from being a multihomer, and finally the fixed fees on the fixed network as well as a mobile network. Note that the type dependent utility from subscribing to mobile services for multihoming consumers $\hat{g}(y)$ may differ from the benefit of singlehoming in a mobile network. In the same way as for singlehomers, the willingness to pay for mobility is an increasing function of consumer type; $\hat{g}'(y) > 0$.

Call demand functions for a subscriber multihoming in mobile network *i* and the fixed network are given by:

$$\hat{q}_i = q_i(p_i, a) = -\frac{\partial \hat{\omega}(p_i, a)}{\partial p_i}, \quad \frac{\partial \hat{q}_i}{\partial p_i} < 0$$

$$\hat{q}_f = q_f(a, p_i) \qquad , \quad \frac{\partial \hat{q}_f}{\partial a} \leq 0$$

Finally, we assume that traffic originated in fixed and traffic originated in mobile are substitutes, i.e.:

$$\frac{\partial q_f}{\partial p_i} \geq 0$$

Note that this assumption *per se* not is contradictory to fixed and mobile services being complements at an aggregate level. Consider the following example, in an uncovered market, a reduction in the fixed usage price will have two opposing effects; a direct substitution effect and an indirect network effect. The indirect network effect is due to some unsubscribing consumers joining the fixed network. This will again result in more potential communication partners, resulting in increased mobile usage. The aggregate effect may be that the network effect dominates the substitution effect such that fixed and mobile services appear to be complements.

3.2. Timing of the games

In this paper we endogenize the homing decisions made by subscribers, i.e. the choice between:

a. Singlehoming in mobile

b. Multihoming in mobile and fixed
c. Singlehoming in fixed

In order to simplify the modelling we will assume that this homing decision is made prior to consumers learning their preferences over mobile services (the *x* parameter, location on the Hotelling line). By doing so the strategic interaction between the two mobile firms will be directly comparable to the reference model.[10]

This timing structure is introduced in order to simplify the modelling, but it can be motivated by assuming that there is a search cost related to learning the characteristics of the mobile services. Suppose consumers only are willing to incur the cost of learning characteristics of the mobile services after the homing decisions are made; i.e. consumers first make their homing decision, and if the decision is to join a mobile network they start searching for the preferable offering.

We assume that mobile termination rates are determined prior to the game we are analyzing. The termination rates are either a result of negotiations between the mobile firms or from regulation. Thus we consider the following multistage game:

1. Consumers make their homing decision
2. Consumers homing in mobile learn their preferences in the mobile dimension, i.e. the location of their preferences on the Hotelling line
3. Mobile firms compete in two-part tariffs

The outcome of stage 1 of the game may be a corner-solution where either all subscribers singlehome in fixed or mobile; alternatively we obtain a corner solution where all consumers multihome. The stage 1 outcome may also be an interior solution where some subscribers choose singlehoming in fixed or mobile, others choose multihoming. The game is solved by backward induction. Thus, in principle, one has to consider all possible stage 1 outcomes. If all subscribers singlehome in mobile we are however back to the reference model. Furthermore, if all subscribers singlehome in fixed, the effect of termination rates vanishes since all traffic will be internal in the fixed network. These outcomes are not interesting in our context. Our focus is the implications of fixed mobile substitution and it turns out that it is sufficient to analyze two outcomes from stage 1 of the game: 1) The outcome where all consumers multihome in fixed and mobile, and 2) the

[10] *Without this assumption consumers located in the middle of the Hotelling line would be more likely to choose the fixed network. Then the strategic interaction in the standard Hotelling model changes. In particular, the change in market share as a result of changing prices takes a different (and more complicated) form.*

outcome characterized by singlehoming, where subscribers are either on fixed or on mobile. In appendix B we also look into two other stage 1 outcomes, namely: B.1.) “An emerging market” where all subscribers are on fixed and some subscribers multihome and B.2.) “A mature market” where all subscribers are on mobile, and some multihome in fixed and mobile.

3.3. The homing decision

The offered mobile services are located on the extremes of the unit line. Consumer preferences are uniformly distributed; thus expected traveling distance is 0.25. The expected disutility from not consuming the most preferred variety is accordingly $\frac{1}{2\sigma}\frac{1}{4}=\frac{1}{8\sigma}$.

Thus at stage 1 of the game a subscriber will choose to singlehome in fixed if this homing decision is preferred over both singlehoming in mobile (*i*) and multihoming (*ii*):

i) $\omega(p_f)-T_f \geq \omega(p_i)-\frac{1}{8\sigma}+g(y)-T_i$

ii) $\omega(p_f) \geq \hat{\omega}(p_i,a)-\frac{1}{8\sigma}+\hat{g}(y)-T_i$

Similarly, a consumer will, at stage 1, prefer to singlehome in mobile over both singlehoming in fixed (*i*) and multihoming (*ii*):

i) $\omega(p_i)-\frac{1}{8\sigma}+g(y)-T_i \geq \omega(p_f)-T_f$

ii) $\omega(p_i)+g(y) \geq \hat{\omega}(p_i,a)+\hat{g}(y)-T_f$

Finally, a consumer will prefer to multihome if:

i) $\hat{\omega}(p_i,a)-\frac{1}{8\sigma}+\hat{g}(y)-T_i \geq \omega(p_f)$

ii) $\hat{\omega}(p_i,a)+\hat{g}(y)-T_f \geq \omega(p_i)+g(y)$

Note that depending upon prices, the shape and locus of the indirect utility functions for singlehoming and multihoming consumers $\omega,\hat{\omega}$ as well as the shape and locus of the additional utility from mobility $g,\hat{g}$, we may end up in scenarios where all subscribers make the same homing decisions or we may end up in mixed situations. As indicated above we will focus on two outcomes: 1) The outcome where all subscribers multihome, and 2) the outcome where some subscribers singlehome in mobile and others singlehome in fixed. In the appendix we briefly look into other outcomes as well.

4. All subscribers multihome

In this section of the paper we will make the extreme assumption that the outcome of stage 1 of the game is that all subscribers choose to multihome. This is the case if at stage 1 of the game, all subscribers prefer multihoming over singlehoming in mobile; i.e. for all $y \in [0,1]$, $\hat{\omega}(p_i, a) + \hat{g}(y) - T_f \geq \omega(p_i) + g(y)$ and they also prefer multihoming over singlehoming in fixed; i.e. for all $y \in [0,1]$, $\hat{\omega}(p_i, a) - \frac{1}{8\sigma} + \hat{g}(y) - T_i \geq \omega(p_f)$.

4.1. Market shares

Market shares of the two mobile firms are determined in the standard Hotelling way:

$$\alpha_i = \frac{1}{2} + \sigma\left(\hat{\omega}(p_i, a) - \hat{\omega}(p_j, a) - T_i + T_j\right)$$

4.2. Stage 3

In the following we will without loss of generality focus on mobile firm 1, and to save notation we write $\alpha = \alpha_1$. Retail profits are:

$$\pi_R = \alpha\left[T_1 + \hat{q}(p_1, a)(p_1 - c - (a - c_0)(1 - \alpha))\right]$$

i.e. market share multiplied with the fixed fee plus profits on traffic. Note that we have made one important simplification in this section; when consumers originate a call in a mobile network, their call will also be terminated in a mobile network.[11]

Consider then profits in the wholesale market. It consists of three elements:

1. Calls from mobile network 2 terminated in network 1: $(a - c_0)\alpha(1 - \alpha)\hat{q}(p_2, a)$

---

[11] *Since all subscribers multihome, all subscribers have a mobile phone, and since we assume that there is no price discrimination between traffic terminated in fixed and mobile, subscribers will (weakly) prefer to terminate calls on mobile phones.*

2. Calls from multihoming subscribers in network 1 originated in fixed, terminating in mobile network 1:
   $(a-c_0)\alpha_1{}^2\hat{q}_f(p_1,a)$
3. Finally calls from multihoming subscribers in network 2 originated in fixed, terminating in mobile network 1
   $(a-c_0)\alpha_1(1-\alpha_1)\hat{q}_f(p_2,a)$

We substitute $T_1=\hat{\omega}(p_1,a)-V_1$, collect terms and obtain the following profit function:

$$\pi_1=\max_{p_1,V_1}\begin{bmatrix}\alpha\left(\hat{\omega}(p_1,a)-V_1+\hat{q}(p_1,a)(p_1-c-(a-c_0)(1-\alpha))\right)\\ +(a-c_0)\left(\alpha(1-\alpha)\hat{q}(p_2,a)+\alpha^2\hat{q}_f(p_1,a)+\alpha(1-\alpha)\hat{q}_f(p_2,a)\right)\end{bmatrix}$$

This function is to be maximized with respect to net utility V_1 and usage price p_1. The first order conditions are:

$$\frac{\partial\pi_1}{\partial p_1}=\alpha\frac{\partial\hat{q}(p_1,a)}{\partial p_1}(p_1-c-(a-c_0)(1-\alpha))+\alpha^2(a-c_0)\frac{\partial\hat{q}_f(p_1,a)}{\partial p_1}=0$$

$$\begin{aligned}\frac{\partial\pi_1}{\partial V_1}&=-\alpha+\sigma\left(\hat{\omega}(p_1,a)-V_1+\hat{q}(p_1,a)(p_1-c-(a-c_0)(1-2\alpha))\right)\\&\quad+(a-c_0)\sigma\left((1-2\alpha)\hat{q}(p_2,a)+2\alpha\hat{q}_f(p_1,a)+(1-2\alpha)\hat{q}_f(p_2,a)\right)=0\end{aligned}$$

Consider first optimal usage price:

$$\frac{\partial\pi_1}{\partial p_1}=0\Leftrightarrow p_1=c+(a-c_0)(1-\alpha)-\alpha(a-c_0)\frac{\frac{\partial\hat{q}_f(p_1,a)}{\partial p_1}}{\frac{\partial\hat{q}(p_1,a)}{\partial p_1}}$$

Proposition 1:
As compared to a model without fixed mobile substitution, multihoming and fixed mobile substitution results in:

- An upward adjustment of usage prices if the termination margin is positive.
- A downward adjustment of usage prices if the termination margin is negative.

Proof:

The result follows directly from fixed and mobile being substitutes, $\frac{\partial \hat{q}_f(p_1, a)}{\partial p_1} \geq 0$;

$$\operatorname{sign}\left[-\alpha(a - c_0) \frac{\frac{\partial \hat{q}_f(p_1, a)}{\partial p_1}}{\frac{\partial \hat{q}(p_1, a)}{\partial p_1}} \right] = \operatorname{sign}[(a - c_0)] \text{ QED}$$

This result is in contrast to the result on usage pricing in the benchmark model in section 2 of this paper. Mobile firms deviate from pricing at perceived marginal cost when traffic originated in the fixed network is a substitute for traffic in the mobile network. This adjustment is increasing in the cross price effect and decreasing in the own price effect.

Assume there is a positive termination margin, and take pricing at perceived marginal cost as a starting point, then a marginal increase in the usage price will result in two effects: 1) An increase in wholesale profits since consumers will increase the number of calls originated in the fixed network resulting in increased termination revenues. 2) A loss in retail profits since the subscription fee will have to be reduced in order to compensate for the loss in consumer surplus due to the increased usage price. The wholesale effect is a first order effect, whereas the effect on retail profits is a second order effect. When there is a termination margin the mobile firms will accordingly increase their profits by raising usage prices above the perceived marginal cost. A negative termination margin will result in the opposite adjustment in usage prices.

Define:

$$\delta \equiv -\frac{\frac{\partial \hat{q}_f(p_1, a)}{\partial p_1}}{\frac{\partial \hat{q}(p_1, a)}{\partial p_1}} > 0$$

In order to simplify the modelling, we will assume that δ is a constant.[12] The condition for optimal usage price can then be written:

[12] *Fulfilled for linear demand functions*

$$p_1 = c + (1-\alpha)(a-c_0) + \delta\alpha(a-c_0)$$

Proposition 2:
Reciprocal termination rates are profit neutral under full multihoming.

Proof:
Inserting optimal usage price in the condition for optimal net utility yields:

$$0 = -\alpha + \sigma\left(\hat{\omega}(p_1,a) - V_1 + \hat{q}(p_1,a)\alpha(1+\delta)(a-c_0)\right) \\ + (a-c_0)\sigma\left((1-2\alpha)\hat{q}(p_2,a) + 2\alpha\hat{q}_f(p_1,a) + (1-2\alpha)\hat{q}_f(p_2,a)\right)$$

$$V_1 = -\frac{\alpha}{\sigma} + \hat{\omega}(p_1,a) + \\ (a-c_0)\left(\alpha(1+\delta)\hat{q}(p_1,a) + 2\alpha\hat{q}_f(p_1,a) + (1-2\alpha)\left(\hat{q}(p_2,a) + \hat{q}_f(p_2,a)\right)\right)$$

In any symmetric equilibrium market shares = 0.5, thus we obtain equilibrium net utility:

$$V_1 = -\frac{1}{2\sigma} + \hat{\omega}(p_1,a) + (a-c_0)\left(\tfrac{1}{2}(1+\delta)\hat{q}(p_1,a) + \hat{q}_f(p_1,a)\right)$$

Finally, inserting equilibrium-, market shares, net utility and usage prices (where $p_1 = p_2$) into the profit function yields:

$$\pi_1 = \frac{1}{2}\hat{\omega}(p_1,a) \\ -\frac{1}{2}\left\{-\frac{1}{2\sigma} + \hat{\omega}(p_1,a) + (a-c_0)\left(\tfrac{1}{2}(1+\delta)\hat{q}(p_1,a) + \hat{q}_f(p_1,a)\right)\right\} \\ +\frac{1}{2}\hat{q}(p_1,a)\left(\{c + (1-\alpha)(a-c_0) + \delta\alpha(a-c_0)\} - c - (a-c_0)\frac{1}{2}\right) \\ +(a-c_0)\frac{1}{4}\left(\hat{q}(p_2,a) + \hat{q}_f(p_1,a) + \hat{q}_f(p_2,a)\right) = \frac{1}{4\sigma}$$

QED

Under full multihoming mobile firms cannot increase their profits by using the termination rate as a collusive device. The mechanism driving this result is the same as in the benchmark model considered in section 2. Termination revenues are passed on to consumers. Thus a margin on mobile termination will result in a reduction of the mobile fixed fees. This reduction is exactly equal to the generated profits on termination. These profits are partly from fixed to mobile traffic and partly from incoming mobile to mobile traffic.

5. Singlehoming

In this section we will assume that all subscribers, at stage 1 of the game, have chosen to singlehome; i.e. all subscribers {*x*, *y*} are either characterized by preferring singlehoming in mobile over multihoming, $\hat{\omega}(p_i, a) + \hat{g}(y) - T_f \le \omega(p_i) + g(y)$ or they are characterized by $\hat{\omega}(p_i, a) - \frac{1}{8\sigma} + \hat{g}(y) - T_i \le \omega(p_f)$, i.e. preferring singlehoming in fixed over multihoming.

Thus in this section we assume that at stage 1 of the game a fraction of subscribers *m*, where $m \in (0,1)$ has chosen to singlehome in a mobile network, and a fraction (1 – *m*) has chosen to singlehome in the fixed network.

5.1. Stage 3 of the game

A consumer singlehoming in the fixed network has utility $V_f = \omega_f(p_f) - T_f$, whereas a subscriber singlehoming in mobile network *i* has net utility $V_i = \omega(p_i) - T_i$. The two mobile networks are competing over the *m* customers in the mobile segment. The market share of mobile firm 1 of the mobile segment is accordingly:

$$\alpha = \frac{1}{2} + \sigma(\omega(p_1) - \omega(p_2) - T_1 + T_2)$$

Retail profit of firm 1 is now:

$$\pi_R = m\alpha\left[T_1 + q(p_1)(p_1 - c - (a - c_0)(1-\alpha)m - (a_f - c_0)(1-m))\right]$$

The difference from the formulation in the previous section is that all market shares are scaled by *m*, and we have included a term capturing the cost of traffic terminated in the fixed network, $(a_f - c_0)$ where a_f denotes the regulated termination fee in the fixed network.[13] We assume $a_f \le c_0$, i.e. that the regulated termination fee in the fixed network is no larger than the

[13] *In the previous section we considered only multihoming consumers and then we simplified the modelling by assuming that all mobile originated traffic also terminated in mobile phones. Under singlehoming some traffic has to be terminated in fixed in order to allow for calls to the group of customers singlehoming in fixed.*

cost of terminating calls in the mobile network. Consider then profits in the wholesale market. It consists of two elements:

1. Calls from mobile network 2 terminated in network 1:
   $(a-c_0)m^2\alpha(1-\alpha)q(p_2)$
2. Calls originated in the fixed network terminated in mobile network 1:
   $(a-c_0)m(1-m)\alpha_1 q_0(a)$

As in the previous section we substitute $T_1=\omega(p_1)-V_1$, collect terms and obtain the following profit function:

$$\pi=\max_{p_1,V_1}\begin{bmatrix} m\alpha\big(\omega(p_1)-V_1+q(p_1)\big(p_1-c-(a-c_0)(1-\alpha)m\ -\big(a_f-c_0\big)(1-m)\big)\big) \\ +m\alpha(a-c_0)\big((1-\alpha)mq(p_2)+(1-m)q_0(a)\big) \end{bmatrix}$$

Maximization with respect to usage price and net utility yields:

$$\frac{\partial\pi}{\partial p_1}=m\alpha q'(p_1)\big(p_1-c-(a-c_0)(1-\alpha)m\ -\big(a_f-c_0\big)(1-m)\big)=0$$

$$\begin{aligned}\frac{\partial\pi}{\partial V_1}=&-m\alpha\\ &+m\sigma\big(\omega(p_1)-V_1+q(p_1)\big(p_1-c-(a-c_0)(1-2\alpha)m\ -\big(a_f-c_0\big)(1-m)\big)\big)\\ &+m\sigma(a-c_0)\big((1-2\alpha)mq(p_2)+(1-m)q_0(a)\big)=0\end{aligned}$$

Optimal usage price is accordingly:

$$p_1=c+(1-\alpha)m(a-c_0)+\big(a_f-c_0\big)(1-m)$$

As compared to the reference model the usage price is adjusted to reflect the termination rate on fixed, but the result is similar in the sense that usage is priced at perceived marginal cost. Consider next optimal net utility, where we insert optimal usage price and obtain:

$$V_1=-\frac{\alpha}{\sigma}+\omega(p_1)+(a-c_0)\big(\alpha mq(p_1)+(1-2\alpha)mq(p_2)+(1-m)q_0(a)\big)$$

In a symmetric equilibrium we have $p_1=p_2=p_i$ and $\alpha=\frac{1}{2}$, thus equilibrium net utility is:

$$V_1 = -\frac{1}{2\sigma} + \omega(p_1) + (a - c_0)\left(\tfrac{1}{2} mq(p_1) + (1-m)q_0(a)\right)$$

Proposition 3:
Under singlehoming the profit in the mobile sector is proportional to the number of consumers in the mobile segment.

Proof:
Equilibrium profits are:

$$\begin{aligned}\pi &= m\tfrac{1}{2}\left(\omega(p_1) - \left\{-\frac{1}{2\sigma} + \omega(p_1) + (a-c_0)\left(\tfrac{1}{2} mq(p) + (1-m)q_0(a)\right)\right\}\right) \\ &\quad + m\tfrac{1}{2}(a-c_0)\left(\tfrac{1}{2} mq(p) + (1-m)q_0(a)\right) \\ &= m\frac{1}{4\sigma}\end{aligned}$$

QED

Since profits increase with the size of the mobile segment it is interesting to analyze stage 1 of the game in order to study whether the termination rate has an impact on the homing decisions.

5.2. Stage 1, homing decisions

If joining a mobile network, a consumer of type (*x*, *y*) receives expected utility:

$$V_i - \frac{1}{8\sigma} + g(y)$$

Where $\frac{1}{8\sigma}$ is the expected disutility from not consuming the most preferred mobile variety. The utility if joining the fixed network is given by $V_f(a)$. The size of the mobile segment is accordingly determined by finding the taste parameter *y** so that subscribers are indifferent as to singlehoming in fixed or singlehoming in mobile:

$$V_i - \frac{1}{8\sigma} + g(y^*) - V_f = 0 \Leftrightarrow g(y^*) = V_f - V_i + \frac{1}{8\sigma}$$

It is the consumers with high willingness to pay for mobility that join the mobile segment. Thus the size of the mobile sector is $m = 1 - y^*$. The function *g*() is everywhere increasing, thus we can write the number of customers, at stage 2 of the game, choosing to join the mobile segment, as a function of the difference in offered net utilities:

$$m = m\left(V_i - V_f\right) = 1 - g^{-1}\left(V_f - V_i + \tfrac{1}{8\sigma}\right), \quad m' > 0$$

Recall that the net utility from joining the fixed network is a non increasing function of the termination fee, i.e. $V_0'(a) \leq 0$. The size of the mobile segment is accordingly given by the solution of the following system of equations:

i. $\quad m = m\left(V_i - V_f\right), \qquad m' > 0$

ii. $\quad V_i = -\dfrac{1}{2\sigma} + \omega(p) + (a - c_0)\left(\tfrac{1}{2} mq(p) + (1-m)q_f(a)\right)$

iii. $\quad p_i = c + \tfrac{1}{2} m(a - c_0) + \left(a_f - c_0\right)(1-m)$

Proposition 4
Under singlehoming the termination rate is not profit neutral. Profit in the mobile sector is a function of the size of the mobile sector. The size of the mobile sector is a function of the termination rate. Furthermore:

a) The profit of the mobile firms is increasing in the termination rate in the point where the termination rate is cost based.

b) If mobile firms are free to raise the termination rate, the fixed network may be driven out of the market or there may exist an interior solution.

The proof is in the appendix.

In this scenario, the mobile firms have incentives to raise the termination rate above the welfare maximizing level, the reason being that the profits in the mobile sector are proportional to the size of the mobile sector. Since termination revenues are passed on to mobile consumers, the utility of mobile subscribers increases in the termination rate. Thus an increase in the termination rate will result in a larger mobile sector. If the mobile firms are free to set the termination rate they may drive the fixed network out of the market or they may end up in an interior solution.

The fixed network is not necessarily driven out of the market. This result deserves a comment. Starting from cost based termination rates there is a first order effect when increasing the termination rate resulting in increased termination revenues. At stage 3 of the game, these revenues are passed on to consumers. Thus the size of the mobile sector increases and so do the profits of the mobile firms. As the termination rate increases further there is however some effects that come into play and some of these effects constrain the mobile firm’s ability to increase the difference in offered utility:

- The usage price increase which result in a deadweight loss,
- There is a positive price effect and a negative volume effect; the revenues on a given volume of F2M traffic increase, but the volume decreases
- As the termination rate increases, the number of mobile customers sharing the (possibly) decreasing termination revenues increases

Note that, on the one hand, if there is a strong positive link between the termination rate on the mobile network and the regulated downstream prices charged by the fixed network, then it is more likely that the mobile firms are able to drive the fixed network out of the market. On the other hand, even if downstream prices in the fixed sector are unaffected by the mobile termination rate, the mobile firms will gain from increasing the termination rate above the cost based level.

Dessein (2003) considers a case with heterogeneous consumers, non linear pricing and elastic subscription. Similarly to the results presented here, Dessein finds that the termination rate is not profit neutral. In his model mobile firms prefer a termination rate below costs. This result is in contrast to the result above that the mobile firms prefer a high termination rate. The difference is due to network effects. Consumers joining the mobile networks in Dessein’s model are genuinely new network members. Thus, increasing the number of subscribers in one of the competing mobile networks results in increased utility for all consumers in both networks. This is in contrast to the result from our model of fixed mobile substitution. In our model, an expansion of the mobile sector results in a reduction of the fixed sector. Thus the number of communication partners is constant.

According to proposition 4 above, mobile firms prefer a termination rate above marginal costs. Furthermore, proposition 4 seems to indicate that the mobile firms may set a relatively high termination rate such that the fixed network is driven out of the market. Such a high termination rate may however violate conditions for a shared market equilibrium in the mobile sector. In Laffont Rey and Tirole (1998a) appendix B it is demonstrated that if the termination margins are large and/or there is high substitutability between the networks no equilibrium exists. This result is derived in a model with the same structure as our stage 3 game. If termination rates at stage 1 of the game are determined at such a high level that stage 3 equilibrium breaks down, then our modelling is no longer valid since stage 3 results are derived by assuming the existence of a symmetric equilibrium. It is outside the scope of the current paper to analyze such a game.

6. Conclusions

The implication of the analysis in the current paper is that there is a case for regulating mobile termination rates in the growth phases of mobile telephony, whereas there is less need for regulation in mature markets characterized by a stable size of the mobile sector. This seems to be the opposite of the approach taken by regulators in Europe, where mobile firms were free to set termination rates in the growth phase and where regulation is introduced once markets mature.

These results have been derived by considering fixed mobile substitution in a model of mobile network competition. We have demonstrated that the termination rates are profit neutral if the size of the mobile sector is given. An implication of this result is that the mobile termination rate does not have an impact on profits in the mobile sector if all subscribers multihome. Furthermore, the termination rate is also profit neutral if there is fixed mobile substitution of a type where consumers change status from multihoming in fixed and mobile to a status where they singlehome in mobile. In situations where consumers multihome and there is a positive termination margin, mobile firms will set usage prices above perceived marginal cost.

Furthermore, if fixed mobile substitution results in an increased number of mobile subscribers, then the mobile termination rate will have an impact on profits in the mobile sector. The mechanism behind this result is that profits in the mobile sector are proportional to the size of the mobile sector. The size of the mobile sector is an increasing function of the net utility offered to mobile subscribers. This net utility is increasing in the termination rate because termination revenues are being passed on to consumers due to competition in the mobile sector. Thus the mobile termination rate will have an impact on profitability in the mobile sector if the size of the mobile sector is affected.

In a mixed market situation where the size of the mobile sector is not given and there are some subscribers multihoming, the two effects described above will in combination result in two kinds of market distortions. At stage 1 of the game mobile firms will set termination rates above cost in order to induce more subscribers to join the mobile networks, then at stage 3, due to the termination margin and the existence of multihoming subscribers, mobile firms will have an incentive to raise usage prices above perceived marginal cost in order to make multihomers substitute traffic originated in mobile for traffic originated in fixed because it results in increased termination revenues.

Appendix A, Proof of proposition 4

We have the following system of equations:

i) $$m = m(V_i - V_f) = 1 - g^{-1}\left(V_f - V_i + \tfrac{1}{8\sigma}\right), \quad m' > 0$$

ii) $$V_i = -\frac{1}{2\sigma} + \omega(p) + (a - c_0)\left(\tfrac{1}{2} mq(p) + (1-m)q_f(a)\right)$$

iii) $$p_i = c + \tfrac{1}{2} m(a - c_0) + (a_f - c_0)(1-m)$$

Total differentiation yields:

$$\frac{dm}{da} = m'\left(\frac{dV_i}{da} - \frac{\partial V_f}{\partial a}\right)$$

$$\frac{dV_i}{da} = -q(p_i)\frac{dp_i}{da} + \tfrac{1}{2} mq(p) + (1-m)q_f(a)$$
$$+ (a - c_0)\left(\tfrac{1}{2} q(p)\frac{dm}{da} + \tfrac{1}{2} mq'(p)\frac{dp_i}{da} - \frac{dm}{da} q_f(a) + (1-m)\frac{\partial q_f(a)}{\partial a}\right)$$

$$\frac{dp_i}{da} = \tfrac{1}{2} m + \tfrac{1}{2}(a - c_0)\frac{dm}{da} - (a_f - c_0)\frac{dm}{da}$$

Combining these expressions yields:

(A1)

$$\frac{dm}{da} = \frac{m'\left((1-m)q_f(a) + (a - c_0)\left(\tfrac{1}{4} m^2 q'(p) + (1-m)\frac{\partial q_f(a)}{\partial a}\right) - \frac{\partial V_f}{\partial a}\right)}{1 + m'\left[-(a_f - c_0)q(p_i) - (a - c_0)\left(\tfrac{1}{2} mq'(p)\left(\tfrac{1}{2}(a + c_0) - a_f\right) - q_f(a)\right)\right]}$$

a)

Consider first the point of cost based termination rates, $a = c_0$, then the expression simplifies to:

$$\frac{dm}{da}=\frac{m'\left(\left(1-m\right)q_f\left(a\right)-\overbrace{\frac{\partial V_f}{\partial a}}^{-}\right)}{1+m'\left[-\underbrace{\left(a_f-c_0\right)}_{-}q\left(p_i\right)\right]}>0$$

This proves part a) of the proposition. Note that $a_f<c_0$ is a sufficient, but not necessary condition.

b)

Consider next the denominator in the expression (A1):

$$1+m'\left[-\left(a_f-c_0\right)q\left(p_i\right)-\left(a-c_0\right)\left(\tfrac{1}{2}mq'\left(p\right)\left(\tfrac{1}{2}\left(a+c_0\right)-a_f\right)-q_f\left(a\right)\right)\right]$$

When the termination margin is positive we have $\frac{1}{2}\left(a+c_0\right)>c_0$, furthermore, by assumption, $a_f<c_0$, thus $\left(\frac{1}{2}\left(a+c_0\right)-a_f\right)>0$, the denominator is accordingly positive for positive termination margins. The sign of (A1) is accordingly determined by the numerator:

$$m'\left(\left(1-m\right)q_f\left(a\right)+\left(a-c_0\right)\left(\tfrac{1}{4}m^2q'\left(p\right)+\left(1-m\right)\frac{\partial q_f\left(a\right)}{\partial a}\right)-\frac{\partial V_f}{\partial a}\right)=0$$

$$\left(a-c_0\right)\underbrace{\left(\tfrac{1}{4}m^2q'\left(p\right)+\left(1-m\right)\frac{\partial q_f\left(a\right)}{\partial a}\right)}_{-}\geq\frac{\partial V_f}{\partial a}-\left(1-m\right)q_f\left(a\right)$$

$$\left(a-c_0\right)\leq\frac{\overbrace{\left(1-m\right)q_f\left(a\right)}^{+}-\overbrace{\frac{\partial V_f}{\partial a}}^{+}}{\underbrace{-\tfrac{1}{4}m^2q'\left(p\right)-\left(1-m\right)\frac{\partial q_f\left(a\right)}{\partial a}}_{+}}$$

The right hand side of this expression is always positive, but the inequality may not hold for sufficiently high termination margins, thus there may exist an interior optimum where the mobile sector has its maximum size, and that this size is below 1.

Appendix B

B.1. An emerging market; all in fixed, some multihome F&M

In this section we let m denote the size of the segment multihoming in fixed and mobile, and we let α denote the market share of mobile firm 1 within the multihoming segment.

B.1.1. Stage 3

Retail profit of firm 1 is now:

$$\pi_R = m\alpha\left[\hat{T}_1 + \hat{q}(\hat{p}_1)\left(\hat{p}_1 - c - (a - c_0)(1-\alpha)m - \left(a_f - c_0\right)(1-m)\right)\right]$$

Profits in the wholesale market consist of three elements:

1. Calls from multihomers in mobile network 2 terminated in network 1:
   $(a - c_0)m^2\alpha(1-\alpha)\left(\hat{q}(\hat{p}_2) + q_0(\hat{p}_2)\right)$
2. Calls from multihomers in mobile network 1 originated in the fixed network terminating in mobile network 1:
   $(a - c_0)m^2\alpha^2 q_f(\hat{p}_1)$
3. Calls from singlehomers in fixed:
   $(a - c_0)m(1-m)\alpha\tilde{q}(a)$

Collecting terms and substituting for net utility yields the following profit function:

$$\begin{aligned}\pi = {} & m\alpha\left[\omega(\hat{p}_1) - \hat{V}_1 + \hat{q}(\hat{p}_1)\left(\hat{p}_1 - c - (a - c_0)(1-\alpha)m - \left(a_f - c_0\right)(1-m)\right)\right] \\ & + m\alpha(a - c_0)\left[m(1-\alpha)\left(\hat{q}(\hat{p}_2) + q_f(\hat{p}_2)\right) + m\alpha q_f(\hat{p}_1) + (1-m)\tilde{q}(a)\right]\end{aligned}$$

Then we can maximize profits:

$$\frac{\partial \pi}{\partial \hat{p}_1} = 0 \Leftrightarrow \hat{p}_1 = c + (a - c_0)(1-\alpha)m + \left(a_f - c_0\right)(1-m) + (a - c_0)m\alpha\delta$$

where :

$$\delta = -\frac{q_0{}'(\hat{p}_1)}{\hat{q}'(\hat{p}_1)}$$

This pricing rule is similar to the one we derived under full multihoming. Consider next the condition for optimal net utility where we insert the optimal pricing rule:

$$\frac{\partial \pi}{\partial \hat{V}_1} = 0 \Leftrightarrow$$

$$\hat{V}_1 = -\frac{\alpha}{\sigma} + \omega(\hat{p}_1) + \hat{q}(\hat{p}_1)\left((a - c_0)m\alpha\delta + (a - c_0)\alpha m\right)$$
$$+ (a - c_0)\left(m(1 - 2\alpha)\left(\hat{q}(\hat{p}_2) + q_0(\hat{p}_2)\right) + m2\alpha q_0(\hat{p}_1) + (1 - m)\tilde{q}(a)\right)$$

Inserting equilibrium prices and market shares:

$$\hat{V}_1 = -\frac{1}{2\sigma} + \hat{\omega} + (a - c_0)\left(mq_0 + (1 - m)\tilde{q} + \hat{q}\,\tfrac{1}{2}m(1 + \delta)\right)$$

Finally, inserting all equilibrium values back into the profit function:

$$\begin{aligned}
\pi &= m\tfrac{1}{2}\left(\hat{\omega} - \left(-\frac{1}{2\sigma} + \hat{\omega} + (a - c_0)\left(mq_0 + (1 - m)\tilde{q} + \hat{q}\,\tfrac{1}{2}m(1 + \delta)\right)\right)\right) \\
&\quad + m\tfrac{1}{2}\hat{q}\begin{pmatrix} c + (a - c_0)\tfrac{1}{2}m + (a_f - c_0)(1 - m) + (a - c_0)m\tfrac{1}{2}\delta \\ -c - (a - c_0)\tfrac{1}{2}m - (a_f - c_0)(1 - m) \end{pmatrix} \\
&\quad + m\tfrac{1}{2}(a - c_0)\left[m\tfrac{1}{2}(\hat{q} + q_0) + m\tfrac{1}{2}q_0 + (1 - m)\tilde{q}\right] \\
&= m\frac{1}{4\sigma}
\end{aligned}$$

Profits in the mobile sector are accordingly proportional to the size of the mobile sector. Then we are back at the same structure as the one we considered in section 5 of the paper.

B.1.2. Stage 1

At stage 1 of the game consumers choose between becoming singlehomers in fixed or multihomers in fixed and mobile. By similar reasoning as in section 5, the size of the multihoming segment is given by $m = \hat{m}\left(\hat{V}_i - V_f\right), \quad m' > 0$.

Similarly to the case of singlehoming, the size of the mobile sector is then given by the solution of the following system of equations:

(i) $m = \hat{m}\left(\hat{V}_i - V_f\right), \quad m' > 0$

(ii) $$\hat{V} = -\frac{1}{2\sigma} + \hat{\omega}(p_i) - C + (a - c_0)\left(mq_0(a) + (1-m)\tilde{q}(a) + \tfrac{1}{2}m(1+\delta)\hat{q}(p_i)\right)$$

(iii) $p_i = c + (a - c_0)(1-\alpha)m + \left(a_f - c_0\right)(1-m) + (a - c_0)m\alpha\delta$

Proposition 5
In a market characterized by all consumers being in the fixed network and some consumers multihoming in fixed and mobile:

- The profit neutrality result does not hold.
- The profit of the mobile firms is increasing in the point of cost based termination rates.

Proof
The solutions are given as the solution of the following system of equations (where we have inserted optimal usage price):

(i) $m = m\left(\hat{V}_i - V_f\right), \quad m' > 0$

(ii) $$\hat{V} = -\frac{1}{2\sigma} + \hat{\omega}\left(c + \tfrac{1}{2}(a - c_0)m(1+\delta) + \left(a_f - c_0\right)(1-m)\right) + (a - c_0)\begin{pmatrix} mq_0(a) + (1-m)\tilde{q}(a) \\ + \tfrac{1}{2}m(1+\delta)\hat{q}\left(c + \tfrac{1}{2}(a - c_0)m(1+\delta) + \left(a_f - c_0\right)(1-m)\right) \end{pmatrix}$$

Total differentiation of this system yields:

$$\frac{dm}{da} = m'\left(\frac{d\hat{V}}{da} - \frac{\partial V_f}{\partial a}\right)$$

$$
\begin{aligned}
\frac{d\hat{V}}{da} &= -\hat{q}(\hat{p})\left(\tfrac{1}{2}m(1+\delta)+\tfrac{1}{2}(a-c_0)(1+\delta)\frac{dm}{da}-\left(a_f-c_0\right)\frac{dm}{da}\right)\\
&+mq_0(a)+(1-m)\tilde{q}(a)\\
&+\tfrac{1}{2}m(1+\delta)\hat{q}\left(c+\tfrac{1}{2}(a-c_0)m(1+\delta)+\left(a_f-c_0\right)(1-m)\right)\\
&+(a-c_0)\left(m\frac{\partial q_0(a)}{\partial a}+q_0(a)\frac{dm}{da}+(1-m)\frac{\partial\tilde{q}(a)}{\partial a}-\tilde{q}(a)\frac{dm}{da}\right)\\
&+(a-c_0)\tfrac{1}{2}(1+\delta)\hat{q}\left(c+\tfrac{1}{2}(a-c_0)m(1+\delta)+\left(a_f-c_0\right)(1-m)\right)\frac{dm}{da}\\
&+(a-c_0)\tfrac{1}{2}m(1+\delta)\hat{q}'(\hat{p})\left(\begin{array}{l}\tfrac{1}{2}m(1+\delta)+\tfrac{1}{2}(a-c_0)(1+\delta)\frac{dm}{da}\\ -\left(a_f-c_0\right)\frac{dm}{da}\end{array}\right)
\end{aligned}
$$

Consider now, as a reference point, cost based termination rates, i.e. $a=c_0$, then the second equation simplifies to:

$$
\begin{aligned}
\frac{d\hat{V}}{da} &= -\hat{q}(\hat{p})\left(\tfrac{1}{2}m(1+\delta)-\left(a_f-c_0\right)\frac{dm}{da}\right)\\
&+mq_0(a)+(1-m)\tilde{q}(a)+\tfrac{1}{2}m(1+\delta)\hat{q}\left(c+\left(a_f-c_0\right)(1-m)\right)
\end{aligned}
$$

and we can combine the two expressions to obtain:

$$\frac{dm}{da} = m'\left(\begin{array}{l} -\hat{q}(\hat{p})\left(\tfrac{1}{2}m(1+\delta)-\left(a_f-c_0\right)\dfrac{dm}{da}\right) \\ +mq_0(a)+(1-m)\tilde{q}(a)+\tfrac{1}{2}m(1+\delta)\hat{q} \end{array}\right)$$

$$-m'\frac{\partial V_f}{\partial a}$$

$$\frac{dm}{da}\left(1-m'\left(a_f-c_0\right)\hat{q}(\hat{p})\right)$$

$$= m'\left(-\hat{q}(\hat{p})\left(\tfrac{1}{2}m(1+\delta)\right)+mq_0(a)+(1-m)\tilde{q}(a)+\tfrac{1}{2}m(1+\delta)\hat{q}-\frac{\partial V_f}{\partial a}\right)$$

$$\frac{dm}{da} =$$

$$\frac{m'\left(\tfrac{1}{2}(1+\delta)m\hat{q}(\hat{p})+mq_0(a)+(1-m)\tilde{q}(a)+\tfrac{1}{2}m(1+\delta)\hat{q}-\dfrac{\partial V_f}{\partial a}\right)}{\underbrace{1-m'\underbrace{\left(a_f-c_0\right)}_{-}\hat{q}(\hat{p})}_{+}} > 0$$

This is similar to the expression under singlehoming.

QED

The result is similar to what we found under singlehoming. Thus, if mobile firms are free to set termination rates they can increase profits by increasing the termination rate above costs. Furthermore, if the resulting termination rate is sufficiently high, stage 3 equilibrium will break down.

B.2. A mature market, All in mobile, some multihome FM

In this section we assume that consumers, at stage 3 of the game, are divided into two groups, a segment of singlehomers in mobile and a segment of multihomers. The size of the singlehoming segment is m_s and the multihoming segment is $1-m_s$. Firm *i* offers tariffs targeted at the single- and multihoming segments respectively: $\left\{(T_i, p_i),(\hat{T}_i, \hat{p}_i)\right\}$, thus market shares of firm 1 within the two segments become:

$$\alpha_s = \frac{1}{2} + \sigma\left(\omega(p_i) - \omega(p_i) - T_i + T_i\right)$$

$$\alpha_m = \frac{1}{2} + \sigma\left(\hat{\omega}(\hat{p}_1, a) - \hat{\omega}(\hat{p}_2, a) - \hat{T}_1 + \hat{T}_2\right)$$

We assume that firms are able (and allowed) to condition the offered mobile tariff on whether the subscribers are within the single- or multihoming segment, thus they can do third degree price discrimination.[14]

Retail profit is now:

$$\begin{aligned}\pi_R = {} & m_s\alpha_s\left[T + q(p_1)(p_1 - c - (1-\alpha)(a - c_0))\right] \\ & + (1-m_s)\alpha_m\left[\hat{T} - \hat{q}(\hat{p}_1, a)(\hat{p}_1 - c - (1-\alpha)(a - c_0))\right]\end{aligned}$$

Recall that net utility is given by $V_i = \omega(p_i) - T_i$, and $\hat{V}_i = \omega(\hat{p}_i, a) - \hat{T}_i$, substituting *V* for *T* yields:

$$\begin{aligned}\pi_R = {} & m_s\alpha_s\left[\omega(p_1) - V_1 + q(p_1)(p_1 - c - (1-\alpha)(a - c_0))\right] \\ & + (1-m_s)\alpha_m\left[\hat{\omega}(\hat{p}_1, a) - \hat{V}_1 - \hat{q}(\hat{p}_1, a)(\hat{p}_1 - c - (1-\alpha)(a - c_0))\right]\end{aligned}$$

Wholesale profits consist of four elements, incoming traffic from the other mobile network originated by single- and multihomers respectively, and incoming traffic originated in the fixed network by customers multihoming in the two mobile networks:

$$\begin{aligned}\pi_w = {} & m_s\alpha(a - c_0)\left[(1-\alpha_s)q(p_2)\right] \\ & + (1-m_s)\alpha(a - c_0)\left[(1-\alpha_m)\hat{q}(p_2, a) + \alpha_m q_0(\hat{p}_1, a) + (1-\alpha_m)q_0(p_2, a)\right]\end{aligned}$$

In order to simplify calculations we carry out profit maximization in two steps by first maximizing profits, $\pi_R + \pi_W$, subject to the constraint that $\alpha = m_s\alpha_s + (1 - m_s)\alpha_m$, and then finding optimal total market share.[15] Thus we maximize the Lagrangian (where λ is the Lagrange multiplier):

$$L = \pi_R + \pi_W + \lambda(\alpha - m_s\alpha_s + (1 - m_s)\alpha_m)$$

And then we maximize the Lagrangian with respect to total market share α by applying the envelope theorem.

Proposition 6

In a market characterized by all consumers being in one of the mobile networks and some consumers multihoming in fixed and mobile:

---

[14] *Note that the homing decisions made at stage 1 of the game typically is observable.*

[15] *This approach is due to Hahn, J. H., 2004.*

- The profit neutrality result holds
- The usage price charged from singlehoming consumers is at perceived marginal cost
- The usage price charged from multihoming consumers is adjusted upwards (for positive termination margins)
- The fixed fee charge from singlehoming consumers exceeds the fixed fee charged from multihoming subscribers

Proof

Profits, $\pi_R + \pi_W$, are to be maximized subject to the constraint that $\alpha = m_s \alpha_s + (1 - m_s)\alpha_m$, thus we have the following Lagrangian: (where λ is the Lagrange multiplier):

$$
\begin{aligned}
L = {} & m_s \{\alpha_s(\omega(p_1) - V_1 + q(p_1)(p_1 - c - (1-\alpha)(a - c_0)))\} \\
& + (1 - m_s)\{\alpha_m(\hat{\omega}(\hat{p}_1, a) - \hat{V}_1 + \hat{q}(\hat{p}_1, a)(\hat{p}_1 - c - (1-\alpha)(a - c_0)))\} \\
& + \alpha(a - c_0)(m_s(1-\alpha_s)q(p_2) + (1 - m_s)(1 - \alpha_m)\hat{q}(\hat{p}_2, a)) \\
& + \alpha(a - c_0)(\alpha_m(1 - m_s)q_0(\hat{p}_1, a) + (1 - \alpha_m)(1 - m_s)q_0(\hat{p}_2, a)) \\
& + \lambda(\alpha - m_s\alpha_s - (1 - m_s)\alpha_m)
\end{aligned}
$$

Consider:

$$
\frac{\partial L}{\partial p_1} = m_s\alpha_s(-q(p_1) + q(p_1) + q'(p_1)(p_1 - c - (1-\alpha)(a - c_0)))
$$

$$
\Leftrightarrow p_1 = c + (1-\alpha)(a - c_0)
$$

and:

$$\frac{\partial L}{\partial \hat{p}_1} = (1-m_s)\alpha_m\left(-\hat{q}(\hat{p}_1,a)+\hat{q}(\hat{p}_1,a)+\frac{\partial \hat{q}(\hat{p}_1,a)}{\partial \hat{p}_1}(\hat{p}_1-c-(1-\alpha)(a-c_0))\right)$$

$$+\alpha(a-c_0)\alpha_m(1-m_s)\frac{\partial q_0(\hat{p}_1,a)}{\partial \hat{p}_1,}=0$$

$$(1-m_s)\alpha_m\frac{\partial \hat{q}(\hat{p}_1,a)}{\partial \hat{p}_1}\left[\hat{p}_1-c-(1-\alpha)(a-c_0)+\alpha(a-c_0)\frac{\frac{\partial q_0(\hat{p}_1,a)}{\partial \hat{p}_1,}}{\frac{\partial \hat{q}(\hat{p}_1,a)}{\partial \hat{p}_1}}\right]$$

$$\hat{p}_1=c+(1-\alpha)(a-c_0)-\alpha(a-c_0)\frac{\frac{\partial q_0(\hat{p}_1,a)}{\partial \hat{p}_1,}}{\frac{\partial \hat{q}(\hat{p}_1,a)}{\partial \hat{p}_1}}$$

Simplification: as in earlier sections, assume constant $-\frac{\frac{\partial q_0(\hat{p}_1,a)}{\partial \hat{p}_1,}}{\frac{\partial \hat{q}(\hat{p}_1,a)}{\partial \hat{p}_1}}=\delta$.

Then we can write: $\hat{p}_1=c+(1-\alpha)(a-c_0)+\alpha(a-c_0)\delta$

Consider next:

$$\frac{\partial L}{\partial V_1}=-m_s\alpha_s+\sigma m_s(\omega(p_1)-V_1+q(p_1)(p_1-c-(1-\alpha)(a-c_0)))$$
$$-\alpha\sigma(a-c_0)m_s q(p_2)-\lambda m_s\sigma$$
$$=0$$

Inserting optimal usage price and solving with respect to V_1:

$$\frac{\partial L}{\partial V_1}=-m_s\alpha_s+\sigma m_s(\omega(p_1)-V_1)-\alpha\sigma(a-c_0)m_s q(p_2)-\lambda m_s\sigma=0$$

$$\sigma m_s V_1=-m_s\alpha_s+\sigma m_s(\omega(p_1))-\alpha\sigma(a-c_0)m_s q(p_2)-\lambda m_s\sigma$$

$$V_1=-\frac{\alpha_s}{\sigma}+\omega(p_1)-\alpha(a-c_0)q(p_2)-\lambda$$

Then consider:

$$
\begin{aligned}
\frac{\partial L}{\partial \hat{V}_1} = &-(1-m_s)\alpha_m \\
&+(1-m_s)\sigma\{(\hat{\omega}(\hat{p}_1,a)-\hat{V}_1+\hat{q}(\hat{p}_1,a)(\hat{p}_1-c-(1-\alpha)(a-c_0)))\\
&+\alpha(a-c_0)\sigma(1-m_s)(-\hat{q}(\hat{p}_2,a)+q_0(\hat{p}_1,a)-q_0(\hat{p}_2,a))\\
&-\lambda\sigma(1-m_s)\\
=& \ 0
\end{aligned}
$$

Inserting optimal usage price and solving with respect to $\hat{V}_1$:

$$
\begin{aligned}
0 = &-(1-m_s)\alpha_m+(1-m_s)\sigma\{(\hat{\omega}(\hat{p}_1,a)-\hat{V}_1+\hat{q}(\hat{p}_1,a)\alpha(a-c_0)\delta)\\
&+\alpha(a-c_0)\sigma(1-m_s)(-\hat{q}(\hat{p}_2,a)+q_0(\hat{p}_1,a)-q_0(\hat{p}_2,a))\\
&-\lambda\sigma(1-m_s)\Leftrightarrow
\end{aligned}
$$

$$
\begin{aligned}
\hat{V}_1 = &-\frac{\alpha_m}{\sigma}+\hat{\omega}(\hat{p}_1,a)+\hat{q}(\hat{p}_1,a)\alpha(a-c_0)\delta\\
&+\alpha(a-c_0)(-\hat{q}(\hat{p}_2,a)+q_0(\hat{p}_1,a)-q_0(\hat{p}_2,a))-\lambda
\end{aligned}
$$

Consider then optimal target network size:

$$
\begin{aligned}
L = \ & m_s\{\alpha_s(\omega(p_1)-V_1+q(p_1)(p_1-c-(1-\alpha)(a-c_0)))\}\\
&+(1-m_s)\{\alpha_m(\hat{\omega}(\hat{p}_1,a)-\hat{V}_1+\hat{q}(\hat{p}_1,a)(\hat{p}_1-c-(1-\alpha)(a-c_0)))\\
&+\alpha(a-c_0)(m_s(1-\alpha_s)q(p_2)+(1-m_s)(1-\alpha_m)\hat{q}(\hat{p}_2,a))\\
&+\alpha(a-c_0)(\alpha_m(1-m_s)q_0(\hat{p}_1,a)+(1-\alpha_m)(1-m_s)q_0(\hat{p}_2,a))\\
&+\lambda(\alpha-m_s\alpha_s-(1-m_s)\alpha_m)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha} = \ & m_s\alpha_s q(p_1)(a-c_0)+(1-m_s)\alpha_m\hat{q}(\hat{p}_1,a)(a-c_0)\\
&+(a-c_0)(m_s(1-\alpha_s)q(p_2)+(1-m_s)(1-\alpha_m)\hat{q}(\hat{p}_2,a))\\
&+(a-c_0)(\alpha_m(1-m_s)q_0(\hat{p}_1,a)+(1-\alpha_m)(1-m_s)q_0(\hat{p}_2,a))\\
&+\lambda=0\Leftrightarrow
\end{aligned}
$$

$$
\begin{aligned}
\lambda = &-(a-c_0)[m_s\alpha_s q(p_1)+(1-m_s)\alpha_m\hat{q}(\hat{p}_1,a)]\\
&-(a-c_0)[m_s(1-\alpha_s)q(p_2)+(1-m_s)(1-\alpha_m)\hat{q}(\hat{p}_2,a)]\\
&-(a-c_0)[\alpha_m(1-m_s)q_0(\hat{p}_1,a)+(1-\alpha_m)(1-m_s)q_0(\hat{p}_2,a)]
\end{aligned}
$$

We can now characterize equilibrium by combining optimal pricing, optimal net utilities, and optimal market shares, i.e.

$p = p_1 = p_2$, $\hat{p} = \hat{p}_1 = \hat{p}_2$, $\alpha = \alpha_s = \alpha_m = \frac{1}{2}$. Then the condition for optimal market share simplifies to $\lambda = -(a - c_0)[m_s q + (1 - m_s)\hat{q} + (1 - m_s)q_0]$, and we obtain:

$$
\begin{aligned}
V_1 &= -\frac{\alpha_s}{\sigma} + \omega(p_1) - \alpha(a - c_0)q(p_2) - \lambda \\
&= -\frac{1}{2\sigma} + \omega + (a - c_0) - \tfrac{1}{2}(a - c_0)q - \lambda \\
V_1 &= -\frac{1}{2\sigma} + \omega + (a - c_0)\left(\left(m_s - \tfrac{1}{2}\right)q + (1 - m_s)\hat{q} + (1 - m_s)q_0\right)
\end{aligned}
$$

and:

$$
\begin{aligned}
\hat{V}_1 &= -\frac{\alpha_m}{\sigma} + \hat{\omega}(\hat{p}_1, a) + \hat{q}(\hat{p}_1, a)\alpha(a - c_0)\delta \\
&\quad + \alpha(a - c_0)\left(-\hat{q}(\hat{p}_2, a) + q_0(\hat{p}_1, a) - q_0(\hat{p}_2, a)\right) - \lambda \\
\hat{V}_1 &= -\frac{1}{2\sigma} + \hat{\omega} + \tfrac{1}{2}\hat{q}(a - c_0)\delta + (a - c_0)\left(m_s q + \left(\tfrac{1}{2} - m_s\right)\hat{q} + (1 - m_s)q_0\right)
\end{aligned}
$$

Finally, equilibrium values can be inserted into the profit function:

$$
\begin{aligned}
L &= m_s \tfrac{1}{2}\left(\omega - V_1 + q\left(p_1 - c - (1 - \alpha)(a - c_0)\right)\right) \\
&\quad + (1 - m_s)\tfrac{1}{2}\left(\hat{\omega} - \hat{V}_1 + \hat{q}\left(\hat{p}_1 - c - (1 - \alpha)(a - c_0)\right)\right) \\
&\quad + \tfrac{1}{2}(a - c_0)\left(m_s \tfrac{1}{2} q + (1 - m_s)\tfrac{1}{2}\hat{q} + (1 - m_s)q_0\right) \\
&\quad = m_s \tfrac{1}{2}(\omega - V_1) + (1 - m_s)\tfrac{1}{2}\left(\hat{\omega} - \hat{V}_1 + \tfrac{1}{2}(a - c_0)\delta\hat{q}\right) \\
&\quad + \tfrac{1}{2}(a - c_0)\left(m_s \tfrac{1}{2} q + (1 - m_s)\tfrac{1}{2}\hat{q} + (1 - m_s)q_0\right)
\end{aligned}
$$

Then inserting net utilities:

$$
\begin{aligned}
\pi &= m_s \tfrac{1}{2}\left(\omega - \left(-\frac{1}{2\sigma} + \omega + (a - c_0)\left(\left(m_s - \tfrac{1}{2}\right)q + (1 - m_s)\hat{q} + (1 - m_s)q_0\right)\right)\right) \\
&\quad + (1 - m_s)\tfrac{1}{2}\left(\hat{\omega} + \frac{1}{2\sigma} - \hat{\omega} - \tfrac{1}{2}\hat{q}(a - c_0)\delta + \tfrac{1}{2}(a - c_0)\delta\hat{q}\right) \\
&\quad - (1 - m_s)\tfrac{1}{2}(a - c_0)\left(m_s q + \left(\tfrac{1}{2} - m_s\right)\hat{q} + (1 - m_s)q_0\right) \\
&\quad + \tfrac{1}{2}(a - c_0)\left(m_s \tfrac{1}{2} q + (1 - m_s)\tfrac{1}{2}\hat{q} + (1 - m_s)q_0\right) \\
&\quad = \frac{1}{4\sigma}
\end{aligned}
$$

Thus profit neutrality holds.

Consider next equilibrium fixed fees:

$$T = \omega - V_1 = \omega - \left(-\frac{1}{2\sigma} + \omega + (a - c_0)\left((m_s - \tfrac{1}{2})q + (1 - m_s)\hat{q} + (1 - m_s)q_0\right)\right)$$
$$= \frac{1}{2\sigma} - (a - c_0)\left((m_s - \tfrac{1}{2})q + (1 - m_s)\hat{q} + (1 - m_s)q_0\right)$$

$$\hat{T} = \hat{\omega} - \hat{V}_1$$
$$= \hat{\omega} - \left(-\frac{1}{2\sigma} + \hat{\omega} + \tfrac{1}{2}\hat{q}(a - c_0)\delta + (a - c_0)\left(m_s q + (\tfrac{1}{2} - m_s)\hat{q} + (1 - m_s)q_0\right)\right)$$
$$= \frac{1}{2\sigma} - \tfrac{1}{2}\hat{q}(a - c_0)\delta - (a - c_0)\left(m_s q + (\tfrac{1}{2} - m_s)\hat{q} + (1 - m_s)q_0\right)$$

and note that:

$$T - \hat{T} = \frac{1}{2\sigma} - (a - c_0)\left((m_s - \tfrac{1}{2})q + (1 - m_s)\hat{q} + (1 - m_s)q_0\right)$$
$$- \left(\frac{1}{2\sigma} - \tfrac{1}{2}\hat{q}(a - c_0)\delta - (a - c_0)\left(m_s q + (\tfrac{1}{2} - m_s)\hat{q} + (1 - m_s)q_0\right)\right)$$
$$= (a - c_0)\left((-m_s + \tfrac{1}{2} + m_s)q + (-1 + m_s + \tfrac{1}{2} - m_s)\hat{q} + (-1 + m_s + 1 - m_s)q_0\right)$$
$$+ \tfrac{1}{2}\hat{q}(a - c_0)\delta$$
$$= (a - c_0)\tfrac{1}{2}(q - \hat{q}) + \tfrac{1}{2}\hat{q}(a - c_0)\delta > 0$$

The sign indicated above holds if singlehomers in mobile originate more calls in the mobile network as compared to multihomers.

QED

The results above are derived by assuming third degree price discrimination, but the profit neutrality result is likely to hold under second degree price discrimination as well. Seen from a mobile network, singlehoming and multihoming consumers can be seen as high volume and low volume customers respectively. As demonstrated by Dessein 2003 (two type model) and Hahn 2004 (continuum of types) the profit neutrality holds in models with consumer heterogeneity, as long as the total number of mobile subscribers is given.

Literature

Armstrong, M., 1998, Network interconnection in telecommunications, The Economic journal, 108, (May), 545 – 564.

Armstrong, M., 2002, The theory of access pricing and interconnection, in Cave et. Al. (eds) Handbook of Telecommunication economics, Vol 1, 295-384 (Elsevier science, Amsterdam).

Armstrong, M., 2004, Network interconnection with asymmetric networks and heterogeneous calling patterns, information Economics and policy 16, 375 – 390.

Cadima, N. and P. Barros, 2000. The impact of mobile phone diffusion on fixed-link network, CEPR Discussion Paper 2598, London.

Carter, M. and J. Wright, 2003, Asymmetric Network Interconnection, Review of Industrial Organization, 22, pp 27 – 46.

De Bijl, P. and M. Peitz, 2002, Regulation and Entry into Telecommunications Markets, Cambridge University Press.

De Bijl, P. et al. 2004, Interconnected Networks, TILEC report, December 2004, available at: http://www.tilburguniversity.nl/tilec/publications/reports/networks.pdf .

Dessein, W., 2003, Network competition in nonlinear pricing, Rand Journal of Economics, 34, (2), 593 – 611.

Dessein, W., 2004, Network competition with heterogeneous customers and calling patterns, Information Economics and policy, 16, pp 323 – 345.

Gans JS., SP. King and J. Wright, 2005, Wireless communications, in Cave et. al. (eds) Handbook of Telecommunication economics, Vol II, pp 243-288 (Elsevier science, Amsterdam).

Gans, JS. and SP King, 2001, Using 'bill and keep' to interconnect arrangements to soften network competition, Economic Letters, 71, 413 – 420.

Gruber, H. and F. Verboven, 2001a, The evolution of markets under entry and standards regulation – the case of global mobile telecommunications, International Journal of Industrial Organization, 19, 1189-1212.

Gruber, H. and F. Verboven, 2001b, The diffusion of mobile telecommunications services in the European Union, European Economic Review, 45, 577-588.

Hahn, J. H., 2004, Network competition and interconnection with heterogeneous subscribers, International journal of industrial Organization, 22, 611 – 631.

ITU, 2003, Mobile Overtakes Fixed: Implications For Policy And Regulation, available at: http://www.itu.int/osg/spu/ni/mobileovertakes/Resources/Mobileovertakes_Paper.pdf.

Laffont, J.-J., P. Rey and J. Tirole, 1998a, Network Competition: I. Overview and non-discriminatory pricing, RAND Journal of Economics, 29, (1) 1 – 37.

Laffont, J.-J., P. Rey, J. Tirole, 1998b, Network Competition: II. Price discrimination, RAND Journal of Economics, 29, (1) 38 – 58.

Laffont, J.-J. and J. Tirole, 2000, Competition in telecommunications, The MIT press, Cambridge, Mass.

NPT, 2005, The Norwegian Telecommunications Market – First Half 2005, available at: http://www.npt.no/pt_internet/eng/publications/telecom_statistics/The_Norwegian_Telecommunication_Market_First_Half_2005.pdf.

Peitz, M., 2005, Asymmetric access price regulation in telecommunications markets, European Economic Review, 49, 341 – 358.

Peitz, M., T. Valletti and J. Wright,2004, Competition in telecommunications: an introduction, Information Economics and Policy, 16, 315–321.

Valletti, T. and G. Houpis, 2005, Mobile Termination: What is the “Right Charge?, Journal of Regulatory Economics 28, 3, pp 235 – 258.

Wright, J, 2000, Competition and Termination in Cellular Networks, CRNEC Working paper, University of Auckland, available at: http://www.business.auckland.ac.nz/comwebContent/docs/Economics/cellular-new.pdf.

Wright, J., 2002, Access pricing under competition: An application to cellular Networks, The Journal of industrial economics, vol L. No 3, pp 289 – 315.

Section 3

Network Competition when Costs are Heterogeneous

Network Competition when Costs are Heterogeneous*

by:
Bjørn Hansen
Department of Economics
BI, Norwegian School of Management
N-0442 Oslo
Norway
Email: bjorn.hansen@bi.no

April 2006

Abstract:
In this paper we study network competition when costs differ among interconnected networks. Such cost differences are observed in the mobile sector as well as in fixed networks. In the paper we find that cost based regulation will not result in first best market shares. The low cost firm will be too small in equilibrium. This is partly due to tariff mediated network externalities. This result is in contrast to the standard result in the literature on network competition where one assumes symmetric cost structure. In the present paper, the regulator can induce a first best market equilibrium by combining cost based regulation of termination rates with a tax based on the number of subscribers. If such a tax is not an available instrument, the regulator can improve welfare by granting a termination margin to the low cost firm as compared to cost based regulation.

Keywords: Network competition, cost asymmetries, telecommunications
JEL Classifications: D21, L41, L43, L51,

---

* I would like to thank Christian Riis, Øystein Foros and Kenneth Fjell for helpful comments to earlier versions of this paper.

1. Introduction

The telecommunication industry is deregulated in most countries and is becoming increasingly more competitive. Nevertheless it is expected that some services will still require regulatory scrutiny. Wholesale termination of calls is an example of a market where networks, even in competitive environments, is in a *de facto* monopoly position since they are exclusive providers of termination services to their own customer base. Accordingly regulators have implemented, or are considering implementing price regulation of termination rates. Starting with the papers by Laffont Rey and Tirole (1998a,b) and Armstrong (1998) there is a considerable body of papers analysing network competition under various assumptions. These papers provide guidance for regulators when determining termination rates. The implication of cost heterogeneity is however hardly addressed in the literature.

In the present paper we introduce cost heterogeneity in the, by now standard model of network competition. It is demonstrated that an optimal policy is not characterised by termination rates regulated at marginal costs because the low cost firm becomes too small in equilibrium. This is partly due to the assumed type of competition and partly due to tariff mediated externalities. Consumers choosing to subscribe to the high cost network do not take into account that by doing so the price other consumers have to pay to call them becomes high. By choosing to change subscription from the high cost to the low cost network, the cost of calling that particular subscriber would decrease for all other subscribers. When choosing network, subscribers do not take into account this tariff mediated externality. An optimal regulatory scheme can be implemented by introducing instruments in addition to regulated termination rates. By taxing inefficient firms and/or subsidising efficient firms based on the number of subscribers one obtains optimal market shares. To our knowledge such taxation and subsidisation are not available instruments to regulators. If regulators are restricted to using regulation of termination rates as the only regulatory instruments we demonstrate that the equilibrium market share of the efficient firm increases if this firm is granted a (small) margin on termination. The welfare gain from increased market share of the efficient firm will however have to be balanced against the deadweight loss due to increasing prices above marginal costs. In the paper we demonstrate that the positive market share effect dominates for small termination margins, thus it is welfare improving to grant a small termination margin to the most efficient firm. In special cases it may be optimal to set a reciprocal termination fee equal to the marginal cost of the high cost firm.

There is reason to believe that marginal cost of terminating calls will differ among some types of networks. In the current paper we have two particular cases in mind. The first case is a situation where a traditional fixed telephony network is competing with an IP based telephony network. The other case we have in mind is when two competing mobile networks based on different radio spectrum allocations compete.[1]

IP telephony comes in many varieties, some with characteristics very similar to traditional telephony, seen from the consumer side. By connecting an adapter to any broadband link with a standard interface (Ethernet), the customer can connect any traditional telephone terminal to the adapter. The customer will hear a dial tone, can use traditional phone numbers and reach any other phone (examples of providers of such services are Vonage in the US and Telio in Norway). The cost structure of IP based telephony is different from the cost structure of running a traditional telephony network. This is mainly due to economies of scope. A traditional telephony network is dedicated for one single service, whereas an IP network is multipurpose. The infrastructure cost is accordingly shared among several services. The cost of providing IP telephony to a customer that also consumes other services based on IP is accordingly low compared to the cost of providing traditional telephony. One might argue that since the cost of IP telephony is lower than the cost of providing traditional telephony, IP will rapidly replace traditional telephony. If so, the question of analysing interconnection of networks with different cost structures is only of interest in a transitory phase. The cost advantage is however based on economies of scope between a set of communication services. It is likely that a significant proportion of the customers will only buy one single communication service; telephony. Thus it is likely that both traditional telephony and IP based telephony will have positive market shares and accordingly exist side by side over a period.

Mobile telephony is based on usage of the radio spectrum. The radio spectrum is a scarce resource, and competing mobile networks are typically based on different spectrum allocations resulting in different cost structures. In a European context, mobile networks are based on 900 MHz licenses, on 1800 MHz licences, and/or UMTS licences (a number of frequency blocks in the range 1900 to 2200 MHz). It is likely that this variation results in cost

[1] Another evident example of networks with different cost structure is when fixed and mobile networks interconnect. In the current paper the focus is on competing networks where consumers choose to connect to one, and only one of the networks. This is not necessarily the case for fixed/mobile. A large proportion of the mobile customer base is also subscribers to the fixed network (so-called multihoming).

differences since the geographical area that can be covered by a single radio cell is a function of frequency.

There is some empirical evidence supporting the assertion that there are differences in marginal costs between telephony networks. Correa (2003) estimates a cost function for fixed line telephony providers in the UK and finds significant differences in marginal costs of providing local calls between firms based on traditional fixed line technology as compared to firms based on cable TV technology.[2] The Competition Commission in the UK, based on an engineering model, estimated the difference in long run incremental network cost of termination mobile calls between combined 900/1800 operators and 1800 operators to be in the range 12% -18% (see Competition Commission 2003, table 2.8). Another example is the cost calculations done by the Swedish regulator PTS where they also discovered cost differences (see PTS 2004). They did not however publish the exact cost differences.[3]

Regulators determine termination rates in many countries. As indicated above the literature does not provide much guidance as to how one should deal with heterogeneous costs. Interestingly, the regulators in both Sweden and in the UK identified cost differences in the mobile sector, but they have chosen quite different approaches. In the UK, estimated cost differences are exactly reflected in the regulated termination rates, i.e. the termination rates differ among the networks. In Sweden the regulator chose to set the same termination fee for all the three regulated mobile networks and it was set at the highest estimated level.

The literature on network competition was initiated by Laffont Rey and Tirole (1998a,b) and Armstrong (1998). Introductions to this literature can be found in Laffont and Tirole (2000) as well as in Armstrong (2002), and an overview of some recent contributions is provided in Peitz et al. (2004). In most of these works it is assumed identical cost structure and reciprocal

[2] On the one hand, the number of TV subscribers connected to the network has a significant positive effect on the marginal cost of providing local calls in her study. On the other hand, the number of TV subscribers has a significant negative effect on the cost of providing telephony subscriptions. The implication is that cable TV networks, as compared to traditional telephony networks, can provide telephony subscriptions at lower marginal costs, whereas the marginal costs of providing calls are higher.

[3] It is not a trivial task to measure incremental cost, and it is likely that some fixed and/or common costs are included in the LRIC results referred above. Fixed and common costs are typically not relevant in pricing decisions. Due to the technical reasons listed above it is nevertheless likely that marginal costs also differ between networks.

termination fees. Armstrong (1998) and Laffont et al. (1998a) demonstrate that under uniform pricing, a reciprocal termination fee above costs will serve as a collusive device. Under nonlinear pricing, this result changes. Laffont et al. (1998b) demonstrate that the two networks are indifferent with respect to the termination fee under two part tariffs, whereas Gans and King (2001) find that a reciprocal termination fee below cost will serve as a collusive device when networks are allowed to price discriminate between on- and off-net traffic.

The literature accordingly indicates that reciprocal termination fees may serve as a collusive device in some cases. Furthermore, if the termination fee is determined unilaterally each network has an incentive to raise its termination fee well above marginal cost[4] resulting in a welfare loss (Gans and King 2001).

Regulators have recognised these results and thus they are attempting to regulate termination rates. When networks are symmetric, the advice to regulators from the literature is straightforward; it is optimal to set a reciprocal termination fee equal to marginal cost. The picture is however not as straightforward if networks are asymmetric. Two classes of asymmetries have been studied in the literature:[5] 1) Vertical differentiation between the networks, and 2) asymmetric cost structures.

Vertically differentiated networks are studied in Carter and Wright (2003) where the source of the quality differential is motivated by the asymmetry between an entrant and an incumbent. They consider two part tariffs and reciprocal termination fees. The superior network (the incumbent) will then always prefer a termination fee at marginal cost whereas the newcomer may want a termination fee above costs. The termination fee preferred by the high quality network is also the welfare maximising termination fee. Peitz (2005) considers a model fairly identical to the model considered by Carter and Wright, Peitz however focuses on incentives for newcomers to entry.[6] Granting a termination mark-up to the entrant makes consumers better off, but the total welfare is reduced. Since the profits of the newcomer also

[4] Given that the consumer's willingness to pay does not increase with the number of incoming calls, the profit maximising termination fee is the monopoly price. Introducing willingness to pay for receiving calls will result in a downward correction to this price.

[5] There is another strand of literature considering consumer heterogeneity and unbalanced calling patterns, demonstrating that reciprocal termination fees under two part tariffs is still profit neutral, see e.g. Dessein, (2003).

[6] In De Bijl and Peitz, 2002, the dynamics of asymmetric competition and entry is considered in a number of numerical simulations.

increase, Peitz argues that entry is being stimulated. Peitz (2002) demonstrates that most of these results are also valid under price discrimination between on- and off-net traffic.

Armstrong (2004) introduces asymmetric costs and heterogeneous calling patterns into a model of network competition. In this model it is the low cost network that should be granted a termination mark-up in order to stimulate the low cost network to sign up the welfare maximising number of subscribers. The modelling in this paper is however different from the modelling framework in the other papers cited above. In particular, Armstrong analyses a case where demand is inelastic and where a dominant firm is being regulated in the downstream market, and a number of small firms are price takers (a competitive fringe). By assuming inelastic demand, modelling is simplified, but by assuming that usage is independent of marginal prices, there is no welfare loss from granting termination mark-ups for given market shares. Furthermore, since the dominant firm is regulated in the downstream market, there is limited strategic interaction between the firms.

The contribution of the current paper is to take a cost asymmetry similar to the one considered by Armstrong into the standard Laffont Rey Tirole model where firms offer three part tariffs and compete à la Hotelling in the downstream market. Based on this model we are in a position to investigate welfare properties of some policies with respect to the regulation of termination fees.

The three major results in the current paper are: 1) cost based regulation will not result in first best market shares. The most efficient firm will be too small. 2) Taxation and subsidisation based on the number of subscribers can induce the first outcome. 3) As compared to cost based regulation of termination rates, granting a (small) termination mark up to the most efficient firm results in increased welfare.

The first results are in line with the results derived in Armstrong 2004. The model studied by Armstrong does however not allow distinguishing result 2 from result 3. In the Armstrong model a margin on termination services has the same effect as a subsidy based on the number of subscribers since demand is inelastic. In the current paper we demonstrate that the deadweight loss from regulating a price away from the underlying marginal cost is dominated by a positive market share effect for small termination margins. A regulator can accordingly increase welfare by granting (small) margins to low cost firms. This result lends support to the regulatory approach taken in Sweden where the regulation of termination rates in effect results in margins to the efficient firms.

The paper is organised as follows: In section 2 the model is presented, in section 3 the welfare properties of cost based regulations are considered. In section 4 optimal regulation is derived. In section 5 the effects of granting termination margins to the efficient firm are investigated and finally, in section 6 we conclude.

2. The model

We consider a two-stage game; in the first stage the regulator determines the interconnection fees and in the second stage the two networks compete à la Hotelling. The market is covered, i.e. all consumers are signed up to one of the two networks. Thus prices are not affecting market size, but prices affect market shares and usage. For notational simplicity total market size is normalised to 1. It is important to have in mind that results from models on network competition depend on the contracts offered to consumers. There are four different types of contracts typically being discussed in the literature: uniform pricing, price discrimination, two part tariffs and two part tariffs with price discrimination.[7] In the current paper we consider the most general contract; two part tariffs with price discrimination. Network i (i = 1, 2) offers contracts: $\{F_i, p_i, \hat{p}_i\}$ where F_i is the fixed fee (subscription fee), p_i is the per minute price of calling other subscribers of the same network (on – net price), and $\hat{p}_i$ is the price of calling subscribers of the other network (off – net price).

2.1. Demand and market shares

Let y denote the sum of the value of income and the stand alone value of network subscription.[8] Consumer tastes are assumed to be uniformly distributed over a line of length 1. Given quantity of calls made q, a consumer located at x joining network i has utility:

$$y - t|x - x_i| + u(q).$$

The parameter t is a measure of disutility from not consuming the most preferred brand (travelling cost). Our assumption of a fully covered market is fulfilled given that the utility from making calls on the network is sufficiently high. Define:

[7] In LRT 98a and LRT 98b the basic model is derived in all these four cases.

[8] For notational simplicity the two terms income and value of network subscription are added together since the market is covered and all consumers are connected to a network.

$$v(p) \equiv \max_q u(q) - pq\,.$$

Let α_i denote the market share of network *i*. The net value of being a subscriber of network *i* can then be written:

$$V_i = \alpha_i v(p_i) + (1-\alpha_i) v(\hat{p}_i) - F_i\,.$$

Throughout the paper we will focus on shared market equilibriums. Such equilibriums exist as long as the disutility parameter *t* is sufficiently large and the difference in utility ($V_i - V_j$) is not too large. We will later return to the exact parameter restrictions under the different cases considered below. Given the existence of a shared equilibrium, market shares will be determined by the location of the consumer being indifferent between the two networks:

$$V_i - t\alpha_i = V_j - t(1-\alpha_i) \Leftrightarrow \alpha_i = \frac{1}{2} + \frac{1}{2t}(V_i - V_j)$$

By defining $\sigma = \frac{1}{2t}$, substituting for V_i and V_j and rearranging, market shares can be written:

$$\text{(1.)}\qquad \alpha_i = \frac{\frac{1}{2} + \sigma\left[v(\hat{p}_i) - v(p_j) - F_i + F_j\right]}{1 - \sigma\left[v(p_i) + v(p_j) - v(\hat{p}_i) - v(\hat{p}_j)\right]}$$

2.2. Cost structure

There is a fixed cost for connecting customers *f*. Furthermore the marginal costs of on-net traffic for network *i* is assumed to be c_i. This cost can be decomposed into two parts, origination and termination, each assumed to be 50% of the total cost. The cost is assumed to differ between the two networks. Network *i* is assumed to charge a_i for termination services. In order to simplify notation we define true and perceived marginal cost for off-net traffic:

$$\bar{c} = \frac{1}{2}(c_i + c_j)$$

$$\hat{c}_i = \frac{1}{2}c_i + a_j$$

2.3. Benchmark, welfare-maximising solution

As a reference point we start by deriving the welfare maximising solution, i.e. maximising the welfare function given by:

$$
\begin{aligned}
W = y - f - \frac{1}{4\sigma}&\left(2\alpha_i^{\,2} - 2\alpha_i + 1\right) \\
&+ \alpha_i\left(\alpha_i\left(u(q_1) - q_1 c_1\right) + (1-\alpha_i)\left(u(\hat{q}_1) - \hat{q}_1\bar{c}\right)\right) \\
&+ (1-\alpha_i)\left((1-\alpha_i)\left(u(q_2) - q_2 c_2\right) + \alpha_i\left(u(\hat{q}_2) - \hat{q}_2\bar{c}\right)\right)
\end{aligned}
$$

Recall that the total number of subscribers is normalised to one. The interpretation of the welfare function is then straightforward. For market shares α_i (and 1 - α_i) the third term is average disutility from not consuming the most preferred variety, whereas the last to terms give the difference between generated utility and costs for a given number of calls. This function is to be maximised with respect to $q_1, \hat{q}_1, q_2, \hat{q}_2, \alpha_i$. It is straightforward to see from the expression above, that as long as the function *u*() is increasing and concave, optimal usage is given by:

$$
\begin{aligned}
q_i^{\,*} &= \left\{ q_i(c_i) \,\middle|\, u'_{q_i} = c_i \right\} \\
\hat{q}_i^{\,*} &= \left\{ \hat{q}_i(\bar{c}) \,\middle|\, u'_{q_i} = \bar{c} \right\}
\end{aligned}
$$

Define $v_i = u\left(q_i^{\,*}\right) - c_i q_i^{\,*}$, $\bar{v} = u\left(\hat{q}_i^{\,*}\right) - \bar{c}\hat{q}_i^{\,*}$, then the welfare function can be written:

(2.)

$$
W^*(\alpha_i) = y - f - \frac{1}{4\sigma}\left(2\alpha_i^{\,2} - 2\alpha_i + 1\right) + \alpha_i^{\,2} v_i + (1-\alpha_i)^2 v_j + 2\alpha_i(1-\alpha_i)\bar{v}
$$

Differentiating with respect to market share yields:

$$
\frac{\partial W}{\partial \alpha_i} = -\frac{1}{2\sigma}(2\alpha_i - 1) + 2\alpha_i v_i - 2v_j + 2\alpha_i v_j + 2\bar{v} - 4\alpha_i\bar{v}
$$

An interior solution satisfies:

(3.) $$\frac{\partial W}{\partial \alpha_i} = 0 \Leftrightarrow \alpha^* = \frac{\frac{1}{2} - 2\sigma\left(v_j - \bar{v}\right)}{1 - 2\sigma\left(v_i + v_j - 2\bar{v}\right)}$$

This is an interior solution to the maximisation problem iff $\frac{1}{2\sigma} > 2v_i - 2\bar{v}$, when network i is the low cost network. We will throughout the paper assume that this condition is fulfilled.[9]

2.4. Market equilibrium

The firms will maximise their profits by determining an optimal contract $\{F_i, p_i, \hat{p}_i\}$. The profits of each firm can be written:

$$\begin{aligned} \pi_i &= \alpha_i(F_i - f) \\ &+ \alpha_i^2(p_i - c_i)q(p_i) \\ &+ \alpha_i\alpha_j(\hat{p}_i - \hat{c}_i)q(\hat{p}_i) \\ &+ \alpha_i\alpha_j(a_i - \tfrac{1}{2}c_i)q(\hat{p}_j) \end{aligned}$$

The first line is the profits on subscription, the second line is profits from on-net traffic, the third is profits on off-net traffic, and the last line is the profits in the wholesale market. As demonstrated in LRT 98a it is convenient to consider profit maximisation as if firms offer a net surplus $V_i = \alpha_i v(p_i) + (1-\alpha_i)v(\hat{p}_i) - F_i$, and some usage prices[10], thus the firms solve:

$$\begin{aligned} \max_{V_i, p_i, \hat{p}_i} \big[&\alpha_i(\alpha_i v(p_i) + \alpha_j v(\hat{p}_i) - V_i - f) \\ &+ \alpha_i^2(p_i - c_i)q(p_i) \\ &+ \alpha_i\alpha_j(\hat{p}_i - \hat{c}_i)q(\hat{p}_i) \\ &+ \alpha_i\alpha_j(a_i - \tfrac{1}{2}c_i)q(\hat{p}_j)\big] \end{aligned}$$

Note that, for given net surplus *V*, market shares are independent of usage prices. Recall that $v'(p_i) = -q(p_i)$ and consider the first order conditions for optimal on- and off-net prices:

---

[9] The condition for existence and stability is that the networks are sufficiently differentiated, i.e. σ small, and that the differences in costs are not too large and that the termination margins are not too large. One of the firms will corner the market if either of these conditions are violated. This condition is discussed in appendix A.

[10] Note however that games where net utilities are the strategic variables yield equilibria different from the equilibria one obtains when firms use prices as strategic variables. The result with respect to usage prices is however identical in the two types of games.

$$p_i: \quad \alpha_i^{\,2}\left(v'(p_i)+q(p_i)+(p_i-c_i)q(p_i)\right)=0 \Leftrightarrow p_i=c_i$$
$$\hat{p}_i: \quad \alpha_i\alpha_j\left(v'(\hat{p}_i)+q(\hat{p}_i)+(\hat{p}_i-\hat{c}_i)q(\hat{p}_i)\right)=0 \Leftrightarrow \hat{p}_i=\hat{c}_i$$

This is a well-known result (see LRT 1998a). The firms determine usage prices by maximising the sum of producer and consumer surplus, and then they will extract as much consumer surplus as possible via the fixed fee. Since on-net traffic is always priced at marginal cost, we can save notation by defining: $v_i \equiv v(c_i)$. Let m_i be the margin on termination services defined by: $m_i \equiv a_i - \frac{1}{2}c_i$. Then we can write: $\hat{c}_i = \bar{c} + m_j$ since: $\hat{c}_i = \frac{1}{2}c_i + a_j = \frac{1}{2}c_i + \frac{1}{2}c_j + m_j = \bar{c} + m_j$.

Consider now the optimal fixed fees:

(4.) $\max\limits_{F_i}\left[\alpha_i(F_i - f) + \alpha_i\alpha_j m_i q(\bar{c}+m_i)\right]$

Inserting for $\alpha_j = 1-\alpha_i$ and differentiating yields the following set of first order conditions:

$$0=\alpha_i+\frac{\partial\alpha_i}{\partial F_i}(F_i-f)+\frac{\partial\alpha_i}{\partial F_i}(1-2\alpha_i)m_i q(\bar{c}+m_i)$$

Inserting for market shares, and rearranging yields:

(5.)

$$F_i=\left(\frac{1}{2}+\frac{\sigma m_i q(\bar{c}+m_i)}{2\left(1-\sigma\left(v_i+v_j-\hat{v}_i-\hat{v}_j\right)+\sigma m_i q(\bar{c}+m_i)\right)}\right)F_j$$
$$+\frac{f}{2}+\frac{1}{4\sigma}+\frac{\hat{v}_i-v_j}{2}-\frac{m_i q(\bar{c}+m_i)\left[1+2\sigma\left(f-v_i+\hat{v}_j\right)\right]}{4\left(1-\sigma\left(v_i+v_j-\hat{v}_i-\hat{v}_j\right)+\sigma m_i q(\bar{c}+m_i)\right)}$$

For given termination margins, m_i, the system of first order conditions (5.) is a system of linear equations. Equilibrium will typically exist and be stable for sufficiently differentiated networks with not too large cost asymmetries and not too large termination margins. For each of the special cases considered below we will provide conditions for the existence of a shared market equilibrium as well as conditions for stability.[11]

[11] The best response function above is a generalised version of the response functions in e.g. Gans and King (2001), and Laffont Rey and Tirole (1998b). They provide conditions for existence under cost symmetry.

3. Cost based termination fees

As indicated above the literature suggests that regulating termination services to marginal costs yields a socially optimal outcome.[12] In this section of the paper we will investigate whether this result is valid when costs differ among the two networks.

When termination fees are regulated down to marginal cost, i.e. $m_i = 0$, the best response functions in (5.) simplifies to:

$$F_i = \frac{F_j}{2} + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\bar{v} - v_j}{2}$$

where $\bar{v} \equiv v(\bar{c})$. The slope of the best response functions is the same for the two firms, whereas the intercepts differ. Let firm *i* be the low cost firm. We impose the following parameter restriction[13] in order to obtain an interior solution: $\frac{1}{2\sigma} > \left[\frac{2}{3}v_i + \frac{1}{3}v_j - \bar{v}\right]$. The equilibrium is illustrated below:

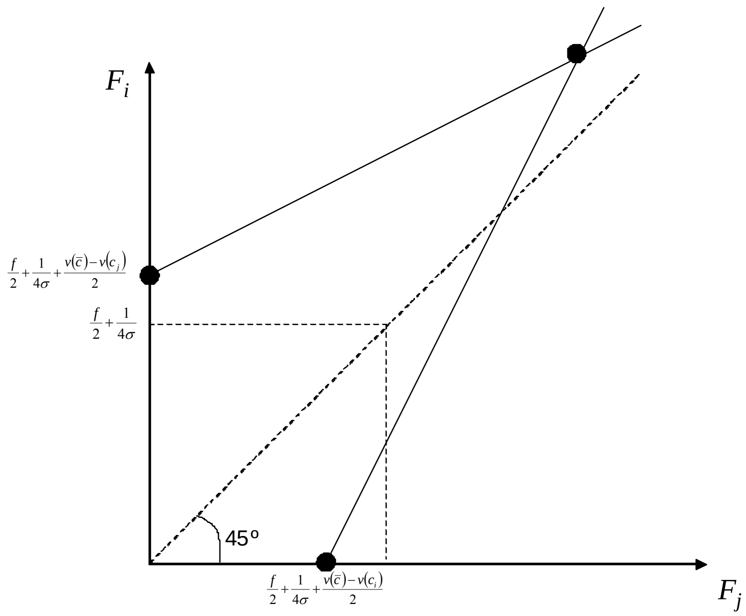

Figure 1, Equilibrium under cost based regulation

---

[12] Even in the Peitz (2005) model, total welfare is reduced when one network (the entrant) is granted a termination mark-up. Peitz however argues that there may be a dynamic gain from allowing termination margins for newcomers because entry is stimulated.

[13] The condition is discussed in appendix B.

From figure 1 we can see directly that the most efficient firm is charging a fixed fee that is higher than the less efficient firm. By combining the two best response functions we can calculate equilibrium fixed fees:

$$F_i = \frac{1}{2\sigma} + f - \frac{1}{3}\left(v_i + 2v_j - 3\bar{v}\right)$$

PROPOSITION 1. *Under cost based regulation, the market share of the most efficient firm is too small compared to the welfare maximising market share.*

PROOF: Let firm *i* be the low cost firm. The equilibrium difference in fixed fees is:

$$F_i - F_j = \frac{1}{3}\left(v_i - v_j\right) > 0 .$$

Consider now the difference in fixed fees that would have induced welfare maximising market shares, ΔF^*. Combining the condition for first best market shares (3.), with the expression for market shares as a function of fixed fees (1.), yields:

$$\frac{\frac{1}{2} - 2\sigma\left(v_j - \bar{v}\right)}{1 - 2\sigma\left(v_i + v_j - 2\bar{v}\right)} = \frac{\frac{1}{2} + \sigma\left[\bar{v} - v_j - \Delta F^*\right]}{1 - \sigma\left[v_i + v_j - 2\bar{v}\right]} \Leftrightarrow$$

$$\Delta F^* = \frac{v_j - v_i}{2\left(1 - 2\sigma\left(v_i + v_j - 2\bar{v}\right)\right)} < 0$$

Thus: $\left(F_i - F_j\right) > 0 > \Delta F^*$. From the market share function (1.) we readily see that the market share of the most efficient firm then becomes too small in equilibrium. QED.

Equilibrium market shares can be calculated by inserting the difference in equilibrium fixed fees, calculated above, into the market share function (1.):

$$\text{(6.)} \quad \alpha_i = \frac{\frac{1}{2} + \sigma\left[\bar{v} - v_j - \frac{v_i - v_j}{3}\right]}{1 - \sigma\left(v_i + v_j - 2\bar{v}\right)} = \frac{3 + 2\sigma\left(3\bar{v} - v_i - 2v_j\right)}{6\left(1 - \sigma\left(v_i + v_j - 2\bar{v}\right)\right)}$$

As demonstrated above, the most efficient firm is too small in equilibrium. This result is partly due to externalities. Since prices differ in equilibrium, the model exhibits tariff mediated network externalities. Consumers on both networks are affected by the network choice made by the indifferent consumer. If one consumer switches from the high cost to the low cost network, the price of making calls to this consumer falls, both for customers in the high cost and customers in the low cost network. Thus, if a consumer switches from the high cost to the low cost network, everybody else is better off. The firms do not however have any incentives to let this externality be reflected in the fixed fees. The two firms compete for the marginal customer taking into consideration the profit contribution from this consumer and without having a mechanism to extract (a fraction of) the increased willingness to pay from all the customers already on the network.

The result above is however only partly due to network externalities. One obtains the same qualitative result in a Hotelling model with differences in marginal costs in the absence of network externalities. The low cost firm does not have incentives to compete sufficiently aggressively for consumers. It becomes too small in equilibrium. The externality effect comes however in addition.

The conditions for having two networks as a welfare maximising solution (A1) and getting a shared market equilibrium (A2) are respectively:

$$\text{(A1)} \quad \frac{1}{2\sigma} > 2v_i - 2\bar{v}$$

$$\text{(A2)} \quad \frac{1}{2\sigma} > \left[\frac{2}{3}v_i + \frac{1}{3}v_j - \bar{v}\right]$$

There exist parameter combinations where the second, but not the first condition is satisfied. Thus we may have equilibrium under cost based regulation where two firms are active, but where a welfare maximising market structure is to only have one network. Following Peitz (2005), one of the networks can be considered as a newcomer. The implication of the result above is then that cost based regulation may stimulate inefficient entry. This case may be relevant in the mobile sector where the licenses to the most cost effective frequencies are allocated first, implying that the last entrant to the market has cost disadvantages relative to the established firms. In the fixed sector we may have the opposite case. Newcomers to the fixed sector are typically based on IP technology which is expected to be more cost efficient. The results above indicate that we may end up in a situation where it would be welfare maximising to switch off the old telephony network, but where market equilibrium yields the opposite result.

4. Optimal regulation

In the section above we demonstrated that cost based regulation of termination fees on the one hand resulted in optimal usage prices, but on the other hand in market shares deviating from the welfare maximising level. If termination rates are altered in order to induce optimal market shares, the result will be that usage prices deviate from the optimal level. The regulator is in a situation where the number of objectives exceeds the number of available instruments. The regulator accordingly needs more instruments in order to induce a welfare maximising outcome.

One obvious alternative for introducing more regulatory instruments is to consider a tax τ_i per subscriber in order to induce a first best market equilibrium. Then profits become:

$$\pi_i = \alpha_i \left(F_i - f - \tau_i\right)$$

and best response functions are:

$$F_i = \frac{F_j}{2} + \frac{f}{2} + \frac{\tau_i}{2} + \frac{1}{4\sigma} + \frac{v(\bar{c}) - v(c_j)}{2}$$

PROPOSITION 2. *The regulator can induce a welfare maximising market equilibrium by setting cost based termination rates and:*

a) *Subsidise the efficient firm per subscriber, or*
b) *Impose a tax per subscriber on the inefficient firm, or*
c) *A combination*

PROOF:
Equilibrium fixed fees are a function of the pair of per subscriber taxes:

$$F_i = f + \frac{\tau_j + 2\tau_i}{3} + \frac{1}{2\sigma} - \frac{1}{3}\left(v_i + 2v_j - 3\bar{v}\right)$$

Optimal taxes must be determined such that they induce welfare maximising market shares, thus they must satisfy:

$$F_i = F_j - \frac{v_i - v_j}{2\left(1 - 2\sigma\left(v_i + v_j - 2\bar{v}\right)\right)}$$

Inserting for equilibrium fixed fees and solving with respect to the tax difference yields:

$$\tau_i - \tau_j = -\left(v_i - v_j\right)\left[1 + \frac{3}{2\left(1 - 2\sigma\left(v_i + v_j - 2\bar{v}\right)\right)}\right] < 0$$

Under our parameter restrictions (A1), the square bracket is positive. Assuming that firm i is the most efficient firm we have $v_i - v_j > 0$. Thus a pair of taxes implementing first best is characterized by $\tau_i - \tau_j < 0$. QED

From the result above we see that it is sufficient to impose a tax on the inefficient firm or to introduce a subsidy to the efficient firm. This result may seem to oppose the regulatory objectives of providing a "level playing field". Instead efficient technologies should be "favoured".

5. Effects of granting a termination margin to the efficient firm

There are, to our knowledge, no examples of regulators having introduced per subscriber taxes and subsidies of the type discussed above. In this section we will consider "second best" regulation, i.e. analyse whether allowing termination margins will result in market equilibrium where market shares are closer to the optimum level.[14]

Firm i is still assumed to be the efficient firm. Assume that the regulator determines a positive termination margin, m, for the efficient firm and applies cost based regulation for the inefficient firm.

In this case usage prices are:

$$p_i = c_i \qquad p_j = c_j$$
$$\hat{p}_i = \bar{c} \qquad \hat{p}_j = \bar{c} + m$$

The best response functions simplify to:

$$F_i = \left(\frac{1}{2} + \frac{\sigma m \hat{q}_j}{2\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right) + \sigma m \hat{q}_j\right)} \right) F_j + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\bar{v} - v_j}{2} - \frac{m \hat{q}_j \left[1 + 2\sigma\left(f - v_i + \hat{v}_j\right)\right]}{4\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right) + \sigma m \hat{q}_j\right)}$$

and:

---

[14] Note that the regulation of Swedish mobile termination can be seen as a special case of such regulation since the regulated reciprocal termination rate is cost based for the least efficient firm. Thus the other two firms, being more efficient, are granted a termination mark-up.

$$F_j = \frac{F_i}{2} + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\hat{v}_j - v_i}{2}$$

By combining the best response functions we obtain equilibrium fixed fees. Explicit expressions are provided in appendix A.3. These equilibrium fixed fees can be inserted into the market share function (1.). Then we obtain market shares as a function of the termination margin:

$$\alpha(m) = \frac{3 - 2\sigma\left(v_i + 2v_j - 2\overline{v}_i - \hat{v}_j\right) + 2m\sigma\hat{q}_j}{6\left(1 - \sigma\left(v_i + v_j - \overline{v}_i - \hat{v}_j\right)\right) + 4m\sigma\hat{q}_j}$$

Under our assumptions there does exist an interior solution for $m = 0$, i.e. under cost based regulation. For sufficiently high termination margins, this may change. Differentiation of the market share function with respect to the termination margin yields:

(7.)

$$\alpha'(m) = 2\sigma \frac{2\sigma\left(-m\frac{\partial\hat{q}_j}{\partial m}\right)\left(v_i + \overline{v} - v_j - \hat{v}_j\right) + \hat{q}_j\left(3 - 2\sigma\left(v_i + 2v_j - 2\overline{v}_i - \hat{v}_j\right) + 2\sigma m\hat{q}_j\right)}{\left(6\left(1 - \sigma\left(v_i + v_j - \overline{v}_i - \hat{v}_j\right)\right) + 4\sigma m\hat{q}_j\right)^2}$$

PROPOSITION 3. *The market share of the efficient firm is monotonously increasing in the termination margin granted to the efficient firm for any non negative termination margin.*

PROOF

A sufficient condition is that the expression in (7.) is strictly positive. The denominator is always positive. A non negative termination margin implies $m \geq 0$. The first bracket in the numerator is then positive (= 0 for $m = 0$) since the demand function has a negative slope. The second bracket is also positive because $v_i > v_j$ by assumption, and $\overline{v} > \hat{v}_j$ for $v_i > v_j$ and $m \geq 0$. Finally, the third bracket is positive by assumption since it is a necessary condition for a shared market equilibrium to exist. QED

The implication of the result above is that the regulator, starting from a cost based equilibrium, can bring the market closer to the optimal market shares by introducing a termination margin to the efficient firm. However, this gain has to be balanced against the deadweight loss resulting from increased usage prices due to the termination margin.

PROPOSITION 4. *Total welfare is increasing in the termination margin at $m = 0$*

The proof of this proposition is provided in the appendix A.4.

By granting a (small) termination margin to the efficient firm, the regulator can accordingly increase welfare.

Note that the regulatory regime in Sweden, where calculated costs differ and where the regulator is imposing a reciprocal termination fee equal to the cost of the least efficient network can be seen as a way of approximating the second best solution of the type we are discussing. Whether the termination margins granted to the more efficient firms in Sweden are too small, exactly equal to, or above the welfare maximising level is however not possible to evaluate based on the model presented here.

The introduction of a termination margin has two effects on the market outcome. A termination margin of the type discussed here results in increased vertical differentiation since the price paid for off-net traffic by customers of the high cost firm increases. We will denote this effect the retail effect. Furthermore, the termination margin has a direct effect on profits in the wholesale market since the low cost firm makes profits on incoming traffic due to the termination margin. In the following we will decompose the effect of introducing a termination margin into the retail effect and a wholesale effect.

Assume that the high cost firm has to pay a margin *m* on outgoing traffic, and assume that the low cost firm is exposed to a tax on incoming traffic such that the regulator exactly confiscates the revenues from the termination margin. Thus the retail effect is present whereas the regulator is cancelling out the wholesale effect due to taxation. By analysing equilibrium in this case we can highlight the effects due to the retail effect.
In this case the best response functions simplify to:

$$F_i = \frac{F_j}{2} + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\overline{v} - v_j}{2}$$

$$F_j = \frac{F_i}{2} + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\hat{v}_j - v_i}{2}$$

Equilibrium fixed fees become:

$$F_i = \frac{1}{2\sigma} + f - \frac{v_i + 2v_j - 2\overline{v} - \hat{v}_j}{3}$$

$$F_j = \frac{1}{2\sigma} + f - \frac{2v_i + v_j - \overline{v} - 2\hat{v}_j}{3}$$

Thus the market share of the low cost firm (firm *i*) becomes:

$$\alpha^r(m) = \frac{3 - 2\sigma\left(v_i + 2v_j - 2\overline{v} - \hat{v}_j\right)}{6\left(1 - \sigma\left(v_i + v_j - \overline{v} - \hat{v}_j\right)\right)}$$

PROPOSITION 5, *As compared to cost based regulation, introducing a termination margin results in the following effects on the market share of the most efficient firm:*

a) A positive retail effect
b) A negative wholesale effect.

Proof:
A sufficient condition for the result above to hold is that the market share of the most efficient firm is larger when we only take the retail effect into consideration as compared to a situation where we consider both effects, i.e. $\alpha(m) < \alpha^r(m)$:

$$\alpha^r(m) - \alpha(m)$$

$$= \frac{3 - 2\sigma\left(v_i + 2v_j - 2\overline{v}_i - \hat{v}_j\right) + 2m\sigma\hat{q}_j}{6\left(1 - \sigma\left(v_i + v_j - \overline{v}_i - \hat{v}_j\right)\right) + 4m\sigma\hat{q}_j} - \frac{3 - 2\sigma\left(v_i + 2v_j - 2\overline{v} - \hat{v}_j\right)}{6\left(1 - \sigma\left(v_i + v_j - \overline{v} - \hat{v}_j\right)\right)}$$

$$= \frac{m\sigma^2\hat{q}_j\left(v_i - v_j + \overline{v} - \hat{v}_j\right)}{3\left(1 - \sigma\left(v_i + v_j - \overline{v} - \hat{v}_j\right)\right)\left(3\left(1 - \sigma\left(v_i + v_j - \overline{v} - \hat{v}_j\right)\right) + 2m\sigma\hat{q}_j\right)} > 0$$

The numerator is positive and the two terms in the denominator are identical to the numerators in the respective market share functions, both positive. The total effect of introducing a termination margin is defined as the sum of the retail and the wholesale effect. We have calculated that the retail effect is larger than the total effect, thus the wholesale effect has to be negative. QED.

The implication of proposition 5 is that the regulator can induce a given market share for the low cost firm at a lower deadweight loss due to distorted prices when the wholesale effect is neutralised. Thus it is welfare improving to introduce taxation in order to extract all net revenues due to the termination margin.

The negative wholesale effect is driven by the fact that the volume of incoming traffic is given by: $\alpha_i(1 - \alpha_i)\hat{q}_j$. For a given termination margin, this volume is maximised for market shares equal to 0.5. We have already demonstrated that market shares for the low cost firm under cost based regulation are characterised by $\alpha_i(0) > 0.5$. Thus, there is an adverse effect for the low cost firm from increasing market shares. In a richer model where we have three or more competing networks and where market shares are below 0.5 for the low cost firm we can expect the wholesale effect to be positive as well (in addition to the positive retail effect). Furthermore, networks typically receive a significant volume of incoming traffic from networks operating in other markets (e.g. incoming traffic from abroad and/or incoming traffic from fixed to a mobile network or *vice versa*). The share of this traffic being received by a network is monotonously increasing in market share. Thus it is likely that the wholesale effect is positive in a richer environment with more than two competing networks and where the competing networks also receive traffic from other markets.

6. Conclusions

In this paper we have studied network competition when costs differ among the interconnected networks. We have analysed the implications of three different principles for regulating termination fees when marginal costs differ. The first case we have analysed is cost based in the sense that termination fees exactly reflect marginal costs. It is a standard result in the literature that marginal prices then are determined at the optimal level. In the current paper we have demonstrated that with cost differences equilibrium market shares are not optimal in this regime. The most efficient network is too small compared to a welfare maximising solution. The reason is partly that with cost differences there is a tariff mediated network externality. There is however no mechanism in the market that enables the efficient firm to internalise this effect.

In the second regulatory regime we introduce taxation and subsidisation, of the two firms based on the number of subscribers as an addition to the cost based regulation of termination rates. By subsidising the low cost firm and/or imposing a tax on the high cost firm, the regulator can implement first best.

In the third regime we investigate whether granting a termination mark-up to the low cost firm can improve the situation as compared to cost based regulation. We have demonstrated that the mark-up has the desired effect on market shares; the low cost firm becomes bigger. Furthermore, we have demonstrated that, starting from cost based regulation, welfare increases as a termination mark-up granted to the low cost firm is introduced. Thus it is welfare improving to let the efficient firm enjoy a (small) mark-up. The effect of granting a termination mark-up to the low cost firm can be decomposed into two parts, a retail effect and a wholesale effect. The retail effect is positive whereas the wholesale effect is negative. This is partly due to the fact that the volume of terminating traffic, where the firm enjoys a mark-up, is maximised for market shares equal to 0.5. Thus wholesale revenues decrease as market shares increase.

7. Appendix

A.1. Conditions for interior solution to the welfare maximisation

The solution:

$$\frac{\partial W}{\partial \alpha} = 0 \Leftrightarrow \alpha^* = \frac{\frac{1}{2} - 2\sigma\left(v_j - \overline{v}\right)}{1 - 2\sigma\left(v_i + v_j - 2\overline{v}\right)}$$

is an optimum if second order conditions are fulfilled and $\alpha^* \in [0,1]$. Consider first the second order conditions:

$$\frac{\partial^2 W}{\partial \alpha^2} = -\frac{1}{\sigma} + 2v_i + 2v_j - 4\overline{v} \leq 0 \Leftrightarrow \frac{1}{2\sigma} \geq v_i + v_j - 2\overline{v}$$

The function $w(c) = u(q^*(c)) - c\ q^*(c)$ is decreasing and convex in c. Then $v(c_i) + v(c_j) > 2\ v((c_i + w_j)/2)$. Thus the right hand side is positive and increasing in the cost difference $c_i - c_j$. The second order condition is accordingly fulfilled as long as the two networks are sufficiently differentiated (σ small) and the cost difference is not too large. $\alpha^* \in [0,1]$ if the following two conditions are fulfilled

$$\left.\frac{\partial W}{\partial \alpha}\right|_{\alpha=0} > 0 \Leftrightarrow 0 < \frac{1}{2\sigma} - 2v_j + 2\overline{v} \Leftrightarrow \frac{1}{2\sigma} > 2v_j - 2\overline{v}$$

$$\left.\frac{\partial W}{\partial \alpha}\right|_{\alpha=1} < 0 \Leftrightarrow 0 > -\frac{1}{2\sigma} + 2v_i - 2\overline{v} \Leftrightarrow \frac{1}{2\sigma} > 2v_i - 2\overline{v}$$

Assume, without loss of generality that $c_i < c_j$, then $v_i > \overline{v} > v_j$, the right hand side of the first condition above is then always negative. The first condition is accordingly always fulfilled. The right hand side of the second condition is always positive. If the cost difference is large, then this condition is violated. Finally, note that when this condition is fulfilled the second order condition above is also fulfilled. Thus a necessary and sufficient condition for an interior solution is:

(A.1) $\frac{1}{2\sigma} > 2v_i - 2\overline{v}$

A.2. Conditions for existence and stability under cost based regulation

Note first that the slope of the best response functions is ½, thus if equilibrium exists, it will be stable.

Under cost based regulation equilibrium, market shares are:

$$\alpha_i^{cb} = \frac{\frac{1}{2}+\sigma\left[\bar{v}-v_j-\frac{1}{3}\left(v_i-v_j\right)\right]}{1-\sigma\left[v_i+v_j-2\bar{v}\right]} = \frac{\frac{1}{2}+\sigma\left[v(\bar{c})-\frac{2}{3}v_j-\frac{1}{3}v_i\right]}{1-\sigma\left[v_i+v_j-2\bar{v}\right]}$$

a shared market equilibrium exists provided that $\alpha_i^{CB} \in (0,1)$. Without loss of generality we assume that firm *i* is the most efficient firm. Then we have the following parameter restriction:

$$1 > \frac{\frac{1}{2}+\sigma\left[v(\bar{c})-\frac{2}{3}v(c_j)-\frac{1}{3}v(c_i)\right]}{1-\sigma\left[v(c_i)+v(c_j)-2v(\bar{c})\right]}$$

$$\frac{1}{2\sigma} > \left[\frac{2}{3}v_i+\frac{1}{3}v_j-\bar{v}\right]$$

Consider then the second order conditions for profit maximisation:

$$\frac{\partial^2\pi}{\partial F_i^2} = \frac{\partial\alpha_i}{\partial F_i}+\frac{\partial\alpha_i}{\partial F_i}+\frac{\partial^2\alpha_i}{\partial F_i^2}\left(F_i-f\right)$$

$$= \frac{-2\sigma}{1-\sigma\left(v_i+v_j-2\bar{v}\right)}$$

The sign of this expression is negative when the denominator is positive. This condition can be written: $\frac{1}{2\sigma} > \frac{1}{2}\left(v_i+v_j-2\bar{v}\right)$ and is fulfilled when the condition above is fulfilled.

The binding condition is accordingly:

(A.2. $\frac{1}{2\sigma} > \left[\frac{2}{3}v_i+\frac{1}{3}v_j-\bar{v}\right]$

This condition can be compared to the condition for an interior solution to the problem of welfare maximisation: $\alpha_i^{CB} \in (0,1)$ is fulfilled.

$$\frac{1}{2\sigma} > 2v_i - 2\bar{v}$$

$$\frac{1}{2\sigma} > \left[\frac{2}{3}v_i + \frac{1}{3}v_j - \bar{v}\right] = 2v_i - 2\bar{v} \underbrace{-\left(\frac{4}{3}v_i - \frac{1}{3}v_j + \bar{v}\right)}_{-}$$

This condition is accordingly always satisfied when the condition $\frac{1}{2\sigma} > 2v_i - 2\bar{v}$ is satisfied. Thus if it is welfare maximising to have two networks then there will always exist an interior equilibrium under cost based regulation. Note that the opposite not is true. There exist parameter combinations such that the welfare maximising outcome is to have only one network but where we obtain shared equilibrium under cost based regulation.

A.3. Deriving market shares when the most efficient firm is granted a margin

In this case usage prices are:

$$p_i = c_i \qquad p_j = c_j$$
$$\hat{p}_i = \bar{c} \qquad \hat{p}_j = \bar{c} + m$$

The best response functions simplify to:

$$F_i = \left(\frac{1}{2} + \frac{\sigma m \hat{q}_j}{2\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right) + \sigma m \hat{q}_j\right)}\right) F_j + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\bar{v} - v_j}{2} - \frac{m\hat{q}_j\left[1 + 2\sigma\left(f - v_i + \hat{v}_j\right)\right]}{4\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right) + \sigma m \hat{q}_j\right)}$$

and:

$$F_j = \frac{F_i}{2} + \frac{f}{2} + \frac{1}{4\sigma} + \frac{\hat{v}_j - v_i}{2}$$

Solving the system of equations above yields the following equilibrium fixed fees:

$$F_i = \frac{1}{2\sigma} + f - \frac{\left(v_i + 2v_j - 2\bar{v} - \hat{v}_j\right)\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right)\right) - 2\sigma m \hat{q}_j\left(\bar{v} - v_j\right)}{3\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right)\right) + 2\sigma m \hat{q}_j}$$

and:

$$F_j = \frac{1}{2\sigma} + f - \frac{\left(2v_i + v_j - \bar{v} - 2\hat{v}_j\right)\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right)\right) + \sigma m \hat{q}_j \left(v_i + v_j - \bar{v} - \hat{v}_j\right)}{3\left(1 - \sigma\left(v_i + v_j - \bar{v} - \hat{v}_j\right)\right) + 2\sigma m \hat{q}_j}$$

The equilibrium fixed fees calculated above can be inserted into the market share function (1.). Then we obtain market shares as a function of the termination margin:

$$\alpha(m) = \frac{3 - 2\sigma\left(v_i + 2v_j - 2\bar{v}_i - \hat{v}_j\right) + 2m\sigma\hat{q}_j}{6\left(1 - \sigma\left(v_i + v_j - \bar{v}_i - \hat{v}_j\right)\right) + 4m\sigma\hat{q}_j}$$

The only terms being functions of *m* are *m*, $\hat{v}_j$ and $\hat{q}_j$, thus we can define constants K_1 and K_2: in order to simplify the expression:

$$\begin{aligned}
\alpha(m) &= \frac{3 - 2\sigma\left(v_i + 2v_j - 2\bar{v}_i - \hat{v}_j\right) + 2m\sigma\hat{q}_j}{6\left(1 - \sigma\left(v_i + v_j - \bar{v}_i - \hat{v}_j\right)\right) + 4m\sigma\hat{q}_j} \\
&= \frac{\overbrace{3 - 2\sigma\left(v_i + 2v_j - 2\bar{v}_i\right)}^{K_1} + 2\sigma\left(\hat{v}_j + m\hat{q}_j\right)}{\underbrace{6\left(1 - \sigma\left(v_i + v_j - \bar{v}_i\right)\right)}_{K_2} + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)} \\
&= \frac{K_1 + 2\sigma\left(\hat{v}_j + m\hat{q}_j\right)}{K_2 + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)}
\end{aligned}$$

Differentiation of the market share function with respect to the termination margin yields:

$$\begin{aligned}
\alpha'(m) = &\frac{2\sigma\left(\frac{\partial \hat{v}_j}{\partial m} + \hat{q}_j + m\frac{\partial \hat{q}_j}{\partial m}\right)\left(K_2 + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)\right)}{\left(K_2 + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)\right)^2} \\
&+ \frac{-2\sigma\left(3\frac{\partial \hat{v}_j}{\partial m} + 2\hat{q}_j + 2m\frac{\partial \hat{q}_j}{\partial m}\right)\left(K_1 + 2\sigma\left(\hat{v}_j + m\hat{q}_j\right)\right)}{\left(K_2 + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)\right)^2}
\end{aligned}$$

After rearranging we obtain:

$$\alpha'(m) = 2\sigma \frac{2\sigma\left(-m\frac{\partial \hat{q}_j}{\partial m}\right)\left(v_i + \bar{v} - v_j - \hat{v}_j\right) + \hat{q}_j\left(K_1 + 2\sigma\left(\hat{v}_j + m\hat{q}_j\right)\right)}{\left(K_2 + 2\sigma\left(3\hat{v}_j + 2m\hat{q}_j\right)\right)^2}$$

$$= 2\sigma \frac{2\sigma\left(-m\frac{\partial \hat{q}_j}{\partial m}\right)\left(v_i + \bar{v} - v_j - \hat{v}_j\right)}{\left(6\left(1-\sigma\left(v_i + v_j - \bar{v}_i - \hat{v}_j\right)\right) + 4\sigma m\hat{q}_j\right)^2}$$

$$+ 2\sigma \frac{\hat{q}_j\left(3 - 2\sigma\left(v_i + 2v_j - 2\bar{v}_i - \hat{v}_j\right) + 2\sigma m\hat{q}_j\right)}{\left(6\left(1-\sigma\left(v_i + v_j - \bar{v}_i - \hat{v}_j\right)\right) + 4\sigma m\hat{q}_j\right)^2}$$

A.4. Proof of proposition 4

Total welfare as a function of the termination margin can be written:

$$W(m) =$$

$$v_0 - f - \frac{1}{4\sigma}\left(2\alpha^2 - 2\alpha + 1\right) + \alpha^2 v_i + (1-\alpha)^2 v_j + \alpha(1-\alpha)\left(\bar{v} + \hat{v}_j + m\hat{q}_j\right)$$

We will call this function *the second best welfare* function. In (2.) we have defined first best welfare as a function of market shares as:

$$W^*(\alpha_i) = v_0 - f - \frac{1}{4\sigma}\left(2\alpha_i^2 - 2\alpha_i + 1\right) + \alpha_i^2 v_i + (1-\alpha_i)^2 v_j + 2\alpha(1-\alpha)\bar{v}$$

With our parameter restrictions, this is a concave function with maximum for $\alpha = \alpha^*$. We can substitute for the first best welfare function in the second best function such that:

$$W(m) = W^*(\alpha_i(m)) - \alpha_i(m)(1-\alpha_i(m))\left(\bar{v} - \hat{v}_j - m\hat{q}_j\right)$$

Let $g(m) \equiv \bar{v} - \hat{v}_j - m\hat{q}_j$ denote the per subscriber deadweight loss. Note that:

$$\frac{\partial g}{\partial m} = -m\frac{\partial \hat{q}_j}{\partial m} > 0 \quad \text{for} \quad m > 0.$$

Differentiation of the second best welfare yields:

$$\frac{\partial W}{\partial m} = \frac{\partial W^*}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial m} - \frac{\partial \alpha_i}{\partial m}(1-2\alpha_i)g(m) - \alpha_i(1-\alpha_i)\frac{\partial g}{\partial m}$$

Consider first the derivative at the point where $m = 0$, where $g = 0$ and $g' = 0$. Then the two last terms become zero and we have:

$$\left.\frac{\partial W}{\partial m}\right|_{m=0} = \underbrace{\frac{\partial W^*}{\partial \alpha_i}}_{+}\underbrace{\frac{\partial \alpha_i}{\partial m}}_{+} > 0$$

Thus, in the point where the termination margin is zero, the second best welfare function is an increasing function of the termination margin.

Letting the termination margin increase above zero has several effects. In particular the per subscriber deadweight loss, $g(m)$ increases, whereas the number of subscribers being exposed to the increased price $(\alpha_i(1-\alpha_i))$ decreases due to the fact that the market share of the low cost firm is an increasing function of the termination margin. Thus there is opposing effects on total welfare from increasing the termination margin. Consider the interval where $m > 0$, and smaller than $\tilde{m} = \{m | \alpha(m) = \alpha^*\}$. Then we have the following signs on the various terms in the welfare function:

$$\frac{\partial W}{\partial m} = \underbrace{\underbrace{\frac{\partial \alpha_i}{\partial m}}_{+}\left(\underbrace{\frac{\partial W^*}{\partial \alpha_i}}_{+} - \underbrace{(1-2\alpha_i)}_{-}\underbrace{g(m)}_{+}\right)}_{+}\underbrace{-\alpha_i(1-\alpha_i)\underbrace{\frac{\partial g}{\partial m}}_{+}}_{-}$$

i.e. the sum of one positive and one negative term. We do however know that close to $m = 0$, the positive terms dominate. Numerical simulations indicate that for reasonable parameter values, the negative dead weight loss effect dominates for sufficiently high termination margins.

8. References

Armstrong, M. 1998, Network interconnection in telecommunications, The Economic journal, 108, (May), 545 – 564.

Armstrong, M. 2002, The theory of access pricing and interconnection, in Cave et al. (eds) Handbook of Telecommunication economics, Vol 1, pp 295-384 (Elsevier science, Amsterdam).

Armstrong, M. 2004, Network interconnection with asymmetric networks and heterogeneous calling patterns, information Economics and policy 16, 375 – 390.

Carter, M. and J. Wright, 2003, Asymmetric Network Interconnection, Review of Industrial Organization, 22, 27 – 46.

Competition Commission, 2003, Vodafone, O2, Orange and T-Mobile: Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks, available at: http://www.competition-commission.org.uk/rep_pub/reports/2003/index.htm.

Correa, L, 2003, Natural or Unnatural Monopolies in UK Telecommunications? Queen Mary University, Dept of economics, working paper no 501, Available at: http://www.econ.qmul.ac.uk/papers/doc/wp501.pdf.

De Bijl, P and M. Peitz, 2002, Regulation and Entry into Telecommunications Markets, Cambridge University Press.

Dessein, W 2003, Network competition in nonlinear pricing, Rand Journal of Economics, 34, (2), 593 – 611.

Gans, JS. and SP. King, 2001, Using 'bill and keep' to interconnect arrangements to soften network competition, Economic Letters, 71, 413 – 420.

Laffont, J.-J., P. Rey, J. Tirole, 1998a, Network Competition: I. Overview and non-discriminatory pricing, RAND Journal of Economics, 29, (1) 1 – 37.

Laffont, J.-J., P. Rey, J. Tirole, 1998b, Network Competition: II. Price discrimination, RAND Journal of Economics, 29, (1) 38 – 58.

Laffont, J-J and J. Tirole, 2000, Competition in telecommunications, The MIT press, Cambridge, Mass.

Peitz, M. 2002, Asymmetric access price regulation in telecommunications markets, mimeo, available at http://yoda.eco.auckland.ac.nz/iep/papers/peitz_2002.pdf.

Peitz, M. 2005, Asymmetric access price regulation in telecommunications markets, European Economic Review, 49, 341 – 358.

Peitz, M., T. Valletti and J. Wright, 2004, Competition in telecommunications: an introduction, Information Economics and Policy, 16, 315–321.

PTS, 2004, LRIC prismetod för terminering av röstsamtal i mobilnät, 5th of July, available at: http://www.pts.se/Archive/Documents/SE/Prismetod_MLRIC.pdf.

Section 4

Competition and compatibility among Internet Service Providers

ELSEVIER Information Economics and Policy 13 (2001) 411–425

INFORMATION ECONOMICS AND POLICY

www.elsevier.com/locate/econbase

Competition and compatibility among Internet Service Providers

Øystein Foros[a], Bjørn Hansen[b,*]

[a]*Norwegian School of Economics and Business Administration/Telenor Research and Development, Hellev. 30, N-5045 Bergen, Norway*

[b]*Telenor Research and Development, P.O. Box 83, N-2027 Kjeller, Norway*

Abstract

We consider a two-stage game between two competing Internet Service Providers (ISPs). The firms offer access to the Internet. Access is assumed to be vertically and horizontally differentiated. Our model exhibits network externalities. In the first stage the two ISPs choose the level of compatibility (i.e. quality of a direct interconnect link between the two networks). In the second stage the two ISPs compete á-la Hotelling. We find that the ISPs can reduce the stage 2 competitive pressure by increasing compatibility due to the network externality. The firms will thus agree upon a high compatibility at stage 1. When it is costly to invest in compatibility, we find that the firms overinvest, as compared to the welfare maximising investment level.

Keywords: Compatibility; Internet; Competition; Duopoly

JEL Classification: L13; L96; L41

1. Introduction

We consider competition between two ISPs (Internet Service Providers) operating in the same geographical area. The product from these service providers is basically access to the Internet, and the ISPs operate their own local network. Internet access is assumed to be horizontally and vertically differentiated from the customer's point of view. Furthermore we assume that there are positive consumption externalities.

*Corresponding author.

E-mail addresses: oystein.foros@nhh.no (O. Foros), bjorn.hansen@telenor.com (B. Hansen).

0167-6245/01/$ – see front matter
PII: S0167-6245(01)00044-0

Competition is modelled as a two-stage game. In the first stage the ISPs determine the quality of interconnection. This choice of interconnection quality can be considered as a choice of compatibility between the networks. In the second stage, for given compatibility, the two firms compete à la Hotelling in attracting customers.

The motivation for the paper is the observation that ISPs competing in the same geographic area typically offer higher quality for on-net communication as compared to off-net communication. Roughly, on-net communication refers to traffic between computers/customers connected to the same ISP, while off-net communication is between computers/customers connected to different networks, e.g. communication between customers subscribing to competing ISPs. Some analysts are arguing that competing ISPs have become more willing to establish private interconnection arrangements. It is however hard to verify this observation because ISPs typically have a non-disclosure policy with respect to the agreements[1].

The majority of the literature on Internet economics focuses on the US-market. In contrast, our paper is motivated by the situation for competing ISPs outside USA. Previously, the attention in the ISP-markets outside USA has been directed to the quality of the connection to the US[2]. The quality of local communication between competing ISPs was rather unimportant since the majority of the Internet content was in the US. The situation is, however, altered, and the portion of the Internet-traffic where both the sender and the receiver are located in the same area is increasing. This tendency is probably due to new customer-types and new services in the Internet. In non-English speaking countries content intended for the mass-market must be produced locally or translated. Furthermore, for new broadband interactive services, such as telemedicine, tele-education, and video conferencing, a relatively larger portion of the communication is probably between customers in the same geographical area as compared to what is the case for conventional Internet services such as web-browsing. Thus, the importance of local interconnection as a strategic variable has increased.

Utility from network participation depends on the number of potential communication partners and the quality of this communication. For given market shares, the customer's willingness to pay is increasing in interconnection quality. It is not obvious, however, that competing firms will choose a high quality. In the presence of network externalities, customers will *ceteris paribus* consider it more advantageous to choose the larger ISP if the chosen quality of interconnection is

[1] The quality in the network is determined by the ratio between capacity and load. The load is varying on a very short time scale. Thus it is hard to observe the quality differential between off and on net traffic from the outside. A customer of a particular ISP will however gain experience over time with particular routes and thus be in a position to assess the quality differential.

[2] Baake and Wichmann (1998) are focusing on the German market, Ergas (2000) and Little et al. (2000) analyze the Australian market, while Mueller et al. (1997) describe the situation in Hong Kong.

reduced. A large ISP may accordingly choose a low interconnection quality in order to increase its market share.

Following several recent studies of the competition in the telecommunication market, e.g. Laffont et al. (1998a,b), we assume that firms offer horizontally differentiated goods. The motivation for this horizontal differentiation is receiving little attention in the literature. In our setting, product differentiation in the horizontal dimension may be given several interpretations. Customers of ISPs are typically buying some complementary products to the Internet access. Private customers connect to the ISP via the telephone line, the television cable or the mobile phone system. Most ISPs are owned by, or, are in co-operation with a supplier of local access, such as cable-TV or local telephony operators. This is one source of horizontal differentiation, since e.g., cable-TV-access suppliers can offer the best incoming capacity, while local telephony companies have more experience with switching technologies and two-way communications. A customer mainly looking for interactive-TV and secondly internet connectivity, will probably prefer the service from an ISP that is a subsidiary of a cable-TV provider. In contrast, for home–office internet connectivity the customer may prefer a subsidiary of a telephone provider. Customers with preferences for mobility choose mobile wireless access although the capacity is lower than for, e.g. cable-TV access.

Another source of horizontal differentiation is the alliances between ISPs and content providers. The ISPs may choose to specialize in offering high quality of some services and thus attracting customers preferring these services. In the AOL–Time Warner merger a hot topic has been whether vertical integration of a content company (Time Warner) and an ISP (AOL) may create incentives to foreclose rivals from accessing some services (“a walled garden strategy”).

1.1. Related literature

There are to our knowledge few papers explicitly considering ISP competition and compatibility choice, but Crémer et al. (2000) and Mason (1999) are notable examples[3].

We are here following Crémer et al. (2000) by modelling network externalities such that customers benefit from an increase in network size, and furthermore, the positive network effect is a function of the degree of compatibility. In contrast to the model in the present paper, Crémer et al. (2000) is assuming that the firms have installed bases and are engaged in Cournot-type competition where the providers compete in attracting new consumers. They find that the firms may have incentives to degrade interconnection quality under market sharing equilibrium.

[3] Other papers looking into ISP competition are DangNguyen and Penard (1999) and Baake and Wichmann (1998). Furthermore, there are some papers looking into congestion control for ISPs under competition, see e.g. Gibbons et al. (2000) and Mason (2000).

This result is in contrast to the results in the present paper and is driven by asymmetries in the installed bases. Thus, in a market with consumer lock-in as in the Cremer et al. model, a large firm may choose a low interconnection quality, whereas in a market with mobile consumers, as in the present paper, a large firm will choose a high interconnection quality.

Mason (1999) models ISP-competition with both horizontal and vertical differentiation, and furthermore, with a timing structure similar to our's. In line with our results, Mason finds that compatibility results in reduced competitive pressure. However, in his paper the firms choose between perfect compatibility and incompatibility at stage 1, and hence, he does not see the positive externality as continuous function of compatibility[4]. Consequently it is not straightforward to consider questions of over-investment in compatibility in the Mason model.

The strategic effect of interconnection quality does also have many similarities with the strategic effect of interconnect price (for given quality) in telephony networks. In telephony networks, the positive externality effect of having many subscribers on competing networks is reduced when the price of making calls across networks increase. In the limiting case with extremely high price of making off-net calls, the telephony subscriber will be indifferent as to the size of the competing network. A high interconnection price will accordingly have similar strategic effects as a low interconnection quality[5]. Both in the telephony interconnection models as well as the present paper, network externalities drives the strategic effect of interconnect quality[6].

Our paper is organized as follows. In Section 2 we present a brief overview of the network structure. In Section 3 we present our model. Finally, in Section 4 we conclude.

2. A brief overview of the network structure

In Fig. 1 we give an illustration of the competition between the ISPs and the choice of compatibility (or interconnection quality). We assume that two ISPs compete in a given market, and we suppose that for communication between own customers (on-net traffic) the ISPs is offering a quality guarantee of $\bar{k}$. If there is

[4] The Mason (1999) model exhibits both vertical and horizontal heterogeneity in consumer preferences. The relative weight of vertical and horizontal aspects is parameterised. In the extreme case with only horizontal consumer heterogeneity, Mason obtains similar result as in the present paper with respect to compatibility.

[5] See Laffont et al. (1998a,b), and Armstrong (1998). Furthermore, in Laffont and Tirole (2000) it is provided an extensive overview of interconnection strategy related to telecommunication.

[6] Such externalities were first given a theoretical treatment by Rohlfs (1974). The strategic effect of network externalities on competition was recognized by Katz and Shapiro (1985). As pointed out by Katz and Shapiro, externalities and the choice of compatibility are closely related.

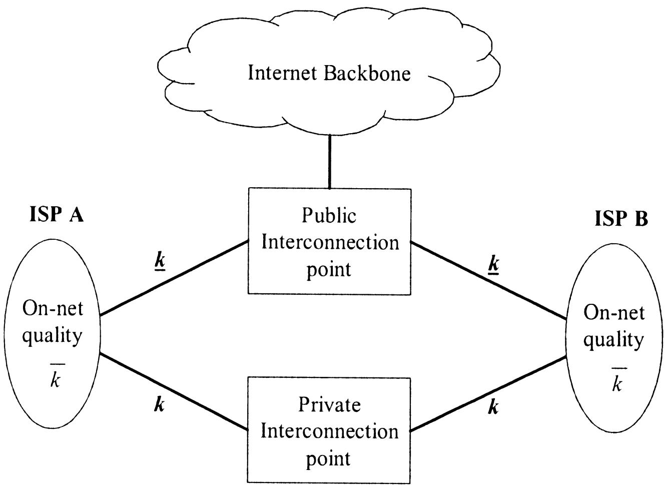

Fig. 1. The interconnection structure.

no private interconnection agreement between the ISPs, no such quality guarantee is given for off-net traffic. Off-net communication between ISP A and ISP B will be sent through a public interconnection point, and the quality level is equal to off-net communication with other destinations in the global Internet Backbone (see Fig. 1). Let the quality of off-net traffic through the public interconnection point be $\underline{k}$ (where $\underline{k} < \bar{k}$). We assume that this public interconnection point is administrated and controlled by a non-commercial third party[7]. The quality level at the public interconnection point is assumed to be outside the control of both ISP A and ISP B[8].

ISP A and ISP B do, however, have the opportunity to invest in a direct link between their networks, i.e. they can invest in a direct interconnect point. If they do, the quality level related to communication between ISP A and ISP B is k, where $\underline{k} \leq k \leq \bar{k}$ (see Fig. 1). The aim of this paper is to analyze the incentives competing ISPs have to implement such direct interconnection. The issue will probably be more important when local access networks are upgraded to high-speed internet communication (broadband) and new bandwidth-demanding services that tolerate minor delays (real time services as interactive video) are offered.

[7] Bailey and McKnight (1997) described four interconnection models where exchange point described here refers to what they called Third-Party Administrator. The other categories are Peer-to-Peer Bilateral, Hierarchical Bilateral, and Co-operative Agreement.

[8] The frequently observed bottleneck problems in public interconnection points in both Europe and the US (see e.g. Kende, 2000) are indications that single ISPs not are able to increase the quality of its services over the public interconnection points. This is probably due to both coordination and free rider problems.

The quality offered in the open internet (the quality level $\underline{k}$) cannot deliver these services[9].

In this paper we will not consider the interplay between the regional ISPs we have in mind and the backbone providers controlling the core global infrastructure (the Internet backbone)[10]. Furthermore, we do not focus on the interplay between ISPs selling internet connectivity and the providers of local access (the last mile into homes). However, as mentioned above, the ISPs and the local access providers are often vertically integrated.

The non-disclosure practice related to private interconnection agreements makes it impossible to know exactly the number of such contracts between competing ISPs[11]. In the US private interconnection agreements are common between the core Internet backbone providers. Also the regional ISPs in Europe have private interconnection agreements with backbone providers at a higher level and in other countries. However, until now, the competing ISPs seem to have been reluctant to implement direct interconnection links in Europe. The non-disclosure characteristic makes it difficult to say whether this trend is changing, but several analysts argue that private interconnection seem to be more common also outside the US[12].

3. The model

The preferences of customers are assumed to be distributed uniformly with density 1 on a line of length 1. The two firms (a and b) are located at the extremes of this unit line, firm a is at $x_a = 0$ and firm b is at $x_b = 1$. The unit cost for each firm is c, and the customers have unit demands. The location of preferences on the unit line indicates the most preferred network type for each customer.

Net utility for a customer located at x connected to supplier i is accordingly:

$$U_i = v_i - t|x - x_i| + \beta \cdot (n_i + kn_j) - p_i \quad \text{where} \quad i, j = a, b \quad i \neq j$$

The first term is a fixed advantage v_i of being connected to network. We define $\theta_i \equiv v_i - v_j$. As long $\theta_i = 0$ there is no vertical differentiation, while the services are vertically differentiated when $\theta_i \neq 0$. The second term is the disutility from not

[9] Interactive services may be among the most profitable services in the Internet. One reason for the profitability of interactive services is that they are less prone to personal arbitrage and reselling than services tolerating some delays (Choi et al., 1997). Another reason is that customers have higher willingness to pay for new information. This will be especially true for strategic information such as stock exchange rates (Shapiro and Varian, 1998).

[10] See Crémer et al. (2000) on the interplay between Internet Backbone Providers.

[11] See Kende (2000) and Gareiss (1999). Kende (2000) gives a comprehensive description of the interconnection agreements between the core backbone providers, and he indicates that as much as 80% of the internet traffic in the US goes through private interconnection points.

[12] See Chinoy and Solo (1997) and Cawley (1997). In Gareiss (1999) there is an overview of private interconnections agreements.

consuming the most preferred network type (the transportation cost in the standard Hotelling model). The third term is a utility term depending upon the number of on-net and off-net customers (n_i and n_j respectively) equal to $\beta \cdot (n_i + kn_j)$, where $\beta \geq 0$ and $k \epsilon [\underline{k}, 1]$. β is measuring the network externality. For $\beta = 0$ consumers are indifferent with the respect to the size of the two networks. The parameter k can be interpreted as a measure of the quality of the interconnect arrangement. When the quality of interconnect equals unity, customers are indifferent as to the distribution of off-net and on-net customers since on- and off-net traffic have identical quality. This is opposed to a situation where $k < 1$. Then, all other things being equal, a customer will prefer a network with many customers. When $k = \underline{k}$ the quality equals the quality available via the Internet (the public interconnection point in Fig. 1), whereas $k > \underline{k}$ implies that the two ISPs have agreed upon establishing an interconnect arrangement (the private interconnection point in Fig. 1) with superior quality. The fourth term, p_i is the per period price charged for ISP subscription[13]. The customers' utility functions are accordingly linear in consumption of the network service and money.

We make the following two assumptions:

Assumption 1. We assume that each of the customers along the interval [0,1] value the products sufficiently high such that they always prefer to subscribe to one or the other network. Thus, the fixed advantage v_i of being connected to either network is sufficiently large.

Assumption 2. There exist one customer in market equilibrium located at x, where $0 < x < 1$, who is indifferent between consuming the network service from the two firms. Thus the valuation differential θ_i between products of the two firms is sufficiently low such that: $|\theta_i| \leq 3(t - \beta(1 - k))$.

We will later demonstrate that Assumption 2 indeed is necessary to obtain a shared market equilibrium. Notice in particular that Assumption 2 implies that $t > \beta(1 - k)$. If this property is violated equilibrium can be characterized by cornering even in ''symmetric'' cases with $\theta_i = 0$ and $p_i = p_j$ because the network externality is dominating the transportation cost[14].

We define α_i as the market share of firm i. Assumptions 1 and 2 are then

[13] Thus, we do not consider any form for usage-based pricing. At first glance, this assumption is more realistic for internet connectivity in the US where flat-rate pricing is the norm for local access. However, we are also observing flat-rate pricing in Europe, in particular for broadband internet connectivity. For a discussion of the usage-based regime in Europe related to Internet access, see e.g. Cave and Crowther (1999).

[14] Assume that almost all customers along the unit line, for some reason, are connected to supplier a. The marginal customer with the longest distance to travel to supplier a, will compare the offer from the two suppliers and he will choose supplier a (and the market will accordingly be characterised by cornering) if: $\beta(1 + \underline{k}0) - t > \beta(0 + \underline{k})$. Thus $t > \beta(1 - \underline{k})$ is ruling out the possibility of market cornering in such symmetric cases.

implying that $n_i = \alpha_i$, $n_j = 1 - \alpha_i$. For a given price vector, the location of preferences $x \in (0,1)$ for the consumer satisfying $U_a = U_b$ is determining the market shares. By defining $\sigma \equiv 1/(2(t - \beta(1-k)))$ we can write the market shares of firm i:

$$\alpha_i = \tfrac{1}{2} + \sigma\theta_i - \sigma(p_i - p_j)$$

σ is a function of k where $\sigma(k) > 0$, $\sigma(1) = 1/2t$, $\sigma'(k) < 0$. Notice that Assumption 2 assures that $\sigma > 0$. The market share functions are very similar to the market share functions in a standard Hotelling model and if $k = 1$ and/or $\beta = 0$, the expression for market shares are identical to what we obtain in a standard Hotelling model with unit demand (i.e. a model without network externalities). In the standard Hotelling model, the parameter σ is interpreted as a measure of product substitutability. The products become closer substitutes if the transportation cost, t, between the two products is reduced. From our definition of σ it also follows that the products become closer substitutes, in the eyes of the consumers, if the quality of the link between the two networks is reduced. We can accordingly expect that an increase in the cost of transport and an increase in the quality of the link between the two networks to have similar effects upon prices and profits.

3.1. The two-stage game

We are considering a two-stage game. In the first stage the two ISPs set the interconnection quality k such that $\underline{k} \leq k \leq 1$. In the second stage, the two ISP simultaneously set their prices for a given k.

3.1.1. Stage 2

In stage 2 the firms set their prices simultaneously, and firm i is choosing p_i so as to maximize profits given by:

$$\pi_i = (p_i - c)\alpha_i = (p_i - c)\left(\frac{1}{2} + \sigma\theta_i - \sigma(p_i - p_j)\right)$$

Combining the first order conditions for firm i and j yields:

$$p_i = \frac{1}{2\sigma} + \frac{\theta_i}{3} + c \quad \text{and} \quad \alpha_i = \frac{1}{2} + \frac{\sigma\theta_i}{3}$$

We will have a shared market equilibrium if and only if $\alpha_i \in (0,1)$ which is satisfied under Assumption 2.

Inserting equilibrium prices and market shares as well as the definition of σ in the profit function and rearranging yields:

$$\pi_i(\theta,k) = \frac{(t-\beta(1-k))}{2} + \frac{\theta_i}{3} + \frac{\theta_i^2}{18(t-\beta(1-k))} \tag{1}$$

When $k=1$ and/or $\beta=0$, this profit function is identical to the one we obtain in a conventional Hotelling model with unit demand.

3.1.2. Stage 1

At stage 1 of the game the two firms decide whether to set up an interconnect arrangement or not. As already stated, stage 2 profit is a function of the quality of interconnection. Direct differentiation of the profit function (1) with respect to k yields:

$$\frac{\partial \pi_i(\theta,k)}{\partial k} = \frac{1}{2}\beta\left(1 - \frac{\theta_i^2}{9(t-\beta(1-k))^2}\right) \tag{2}$$

By definition we have $\theta_j = -\theta_i$, and thus we get:

$$\frac{\partial \pi_i}{\partial k} = \frac{\partial \pi_j}{\partial k}\ \forall k$$

We readily see that the firms do not have conflicting interests with respect to network compatibility, implying that the two firms always agree upon the optimal interconnection quality-level k. Consequently, there is no need for an assumption ensuring that the firm with the lowest incentives for quality has a veto in setting k. The condition for having a shared market equilibrium is $|\theta_i| \leq 3(t-\beta(1-k))$ (Assumption 2). This condition implies that the large bracket above is positive. Thus in any shared market equilibrium profits of both firms increase in interconnect quality.

The effect upon profits from changing interconnect quality can be decomposed into a price and a market share (or volume) effect by differentiating: $\pi_i = \alpha_i(p_i - c)$:

$$\frac{\partial \pi}{\partial k} = \frac{\partial \alpha}{\partial k}(p_i - c) + \alpha_i \frac{\partial p_i}{\partial k}$$

The first term is the market share effect and the second term is the price effect. By inserting the definition of σ in the equilibrium price and differentiating with respect to k we obtain: $\partial p_i / \partial k = \beta$. The price effect is accordingly positive for both firms. This is opposed to the market share effect. When $\theta_i \neq 0$, market shares are functions of interconnect quality. By substituting for σ in the equilibrium market shares and differentiating we obtain:

$$\frac{\partial \alpha_i}{\partial k} = \frac{-\theta_i \beta}{6(t-\beta(1-k))^2} \tag{3}$$

The market share effect is positive for the firm selling the inferior service and thus it is negative for the firm selling the superior service. The negative market share effect for the firm selling the superior product is however dominated by the positive price effect as demonstrated above.

3.1.2.1. Cost free interconnection quality

Assume it is costless to improve the quality of interconnect. As demonstrated above, the differentiated profit function is everywhere increasing in k for both firms. Thus the firms have no incentives to damage the quality of the link between the two networks and furthermore, if possible, they have a mutual interest in improving the quality of this link. Then, both on-net and off-net traffic have the same quality level $k = \bar{k} = 1$.

Prices and profits increasing in the quality of the link between the two networks are due to two effects. First, for given market shares willingness to pay is increasing from all customers as the quality is increased. Second, when the quality of the link is increased the competition between the two suppliers becomes less aggressive[15]. When comparing the conventional Hotelling model with our model featuring network externalities, the argument can be put the other way around: When the networks offer less than perfect connectivity ($k < 1$) then the firms will compete more aggressively than what the conventional Hotelling model predicts.

3.1.2.2. Convex costs of interconnection quality

The assumption above that firms can increase interconnection quality without incurring costs is clearly an unrealistic assumption since both router and transmission capacity is costly in the market place. Furthermore there will be transaction cost of writing a contract and there will typically be costs of mutual monitoring. We can thus add realism to our model by taking into account that interconnection is costly. Then the shape of the interconnection cost function will affect the optimal solution. A necessary condition for an interior solution ($k \in (\underline{k}, 1)$) is that the interconnect cost function is convex.

One can argue that it is reasonable to expect the interconnection cost to be convex, since, as interconnect quality increase, the complexity of the contract the two firms can write becomes large. As the quality of interconnect increase, the joint network of the two suppliers become more like a common facility where the firms have ample opportunities of opportunistic behavior. Firms will typically be reluctant to agree upon interconnection unless the contract prohibits opportunistic behavior. In order to observe and verify that the contract indeed is fulfilled, costly mutual monitoring is required.

In the following we will assume the cost of investing in interconnect quality in

[15] The best response functions (''reaction functions'') in stage 2 of the game is: $p_i = R(p_j) = \frac{1}{2}(t - \beta(1 - k) + p_j + \theta_i + c)$. An increase in k will result in parallel shifts outwards for these best response functions and the firms does indeed become less aggressive as the quality of interconnect increase. We can furthermore see $R' = 0.5$, we are thus considering a stable Nash equilibrium.

order to increase the quality of interconnect k above $\underline{k}$ is $I = I(k)$, where $I(\underline{k}) = 0$, $I' > 0$, $I'' > 0$ $\lim_{k\to 1} I(k) = \infty$ and $\lim_{k\to \underline{k}^+} I'(\underline{k}) = 0$. Assume now that the two firms are forming an input joint venture where they equally share the cost of investing in interconnect quality. Each firm will then maximize the stage 2 profit minus the share of the interconnect cost the firm has to pay in stage 1. Thus the two firms will solve identical optimization problems and agree upon a interconnect quality level k^d characterized by:

$$k^d = \arg\max(\pi^i(\theta,k) - \tfrac{1}{2}I(k)).$$

Thus the investment joint venture investment level is characterized by:

$$I'(k) = \beta - \frac{\beta\theta_i^2}{9(t - \beta(1-k))^2}$$

For $\theta_i \neq 0$ the profit functions are convex in k. With our assumptions we have $\pi'(\underline{k}) > I'(\underline{k})$ and $\pi'(1) < I'(1)$. Thus there exist at least one $k \in (\underline{k},1)$ satisfying the first order conditions. For $\theta_i = 0$ there is one and only one k satisfying the first order condition. The second order conditions are satisfied and this solution is indeed optimal. For $\theta \neq 0$ we cannot rule out the possibility that there is more than one k satisfying the first order condition. A sufficient condition for a single unique solution is that the marginal profit curve and the marginal investment curve cross only once. We will in the following assume that the marginal curves cross only once.

We can compare this equilibrium quality level with the socially optimal quality. The first best interconnect quality, k^*, is defined as the quality level that is maximizing customer gross surplus minus total production cost. Consider, for simplicity, the model in the absence of vertical differentiation (i.e. $\theta_i = 0$). First best is then evidently characterized by sharing customers evenly among the two firms since the unit cost of serving customers in the two firms are identical and customers are distributed uniformly on the interval, Then average distance from the most preferred brand is 0.25. Inserting this average distance as well as the optimal market shares in the utility function yields the following welfare function:

$$k^* = \arg\max[v_i - 0.25t + 0.5\beta \cdot (1+k) - c - I(k)].$$

The first best investment level is then characterized by:

$$0.5\beta = I'(k)$$

This is in contrast to the investment level in the input joint venture. In the absence of vertical differentiation the optimal investment level for the input joint venture is: $\beta = I'(k)$. An input joint venture will thus choose a quality level of the interconnect arrangement exceeding the socially optimal level. In the Appendix we demonstrate that we obtain a similar over investment result in the model under vertical differentiation as well. The intuition behind the over investment result is the following: There are two effects leading to the firms' stage 2 profits increasing in interconnect quality: The first effect is that for given market shares willingness to pay is increasing from all customers as the quality is increased. The second effect is that when the quality of the link is increased, the competition between the two suppliers becomes less aggressive. Only the first effect is a social gain. Thus the input joint venture is over-investing in interconnect quality in order to reduce the stage 2 competitive pressure.

4. Conclusion

In this paper, we have considered the incentives for an Internet Service Provider (ISP) to strategically degrade the interconnection quality with the competitors. We have modeled this in a game where two firms choose the quality of interconnection before they compete over market shares á la Hotelling. In the case where there is no vertical differentiation, the firms split the market equally, and they have no incentives to degrade interconnection quality. Moreover, when interconnection is costly the firms will over-invest in interconnection quality as compared to the first best quality level.

We have also demonstrated that if the products from the two firms also are vertically differentiated, then the firm providing the superior product will have the larger market share. When the necessary conditions for a shared market equilibrium is fulfilled, the firms will agree upon the optimal interconnection quality. Furthermore, if interconnection quality is costly, the firms will agree upon a quality of interconnect exceeding the welfare maximizing quality level.

Finally it is not straightforward to compare the model results with the interconnection policy in the market place due to the non-disclosure policy. Representatives in the industry do however make statements indicating that competing ISPs do interconnect in cases where the two firms in question are sufficiently symmetric. Such observations are lending support to the results of the present paper.

Acknowledgements

For helpful discussions and comments, we thank two anonymous referees, the editor, Hans Jarle Kind, Tore Nilssen, Björn Rupp, Jan Yngve Sand and seminar participants at the 2nd Berlin Internet Economics Workshop, at the EARIE 99 conference, and at Foundation for Research in Economics and Business Administration.

Appendix A. Welfare maximizing interconnect investments

Consumers with preferences to the left of some point α join network a. Since the individual transport cost is tx, and the distribution of consumers is uniform along the line, the sum of travelling costs for all consumers joining the networks are $1/2\alpha^2 t$ and $1/2(1-\alpha)^2 t$ for network a and b respectively. In stage 2 of the game the social welfare function is:

$$\begin{aligned} W(k) = \max_{\alpha}[&\alpha\{v_a - \tfrac{1}{2}\alpha t + \beta(\alpha + k(1-\alpha)) - c\} \\ &+ (1-\alpha)\{v_b - \tfrac{1}{2}(1-\alpha)t + \beta(1-\alpha+k\alpha) - c\}] \end{aligned}$$

The welfare maximizing market share $\alpha*$ is thus:

$$\alpha* = \frac{1}{2} + \frac{\theta_a}{2(t - 2\beta(1-k))}$$

It can be shown that the market share of the firm selling the superior product will be to small in market equilibrium as compared to the welfare maximising market share. In special cases, the welfare maximising solution is to let the firm selling the superior product serve the entire market whereas both firms are active in the market equilibrium Notice that this results not is specific to our model featuring network externalities. With the parameter value $\beta = 0$, the model does not exhibit network externalities (and thus there is no effect upon utility by improving interconnect quality). Then the welfare maximising market share is: $\alpha* = 1/2 + \theta_a/2t$ whereas market equilibrium is characterised by: $\alpha* = 1/2 + \theta_a/6t$. Thus the market share of the firm selling the superior product is to small.

The stage 1 socially optimal investment level is:

$$k* = \arg\max(W(k) - I(k))$$

$$\text{FoC: } W' = I'$$

By applying the envelope theorem on $W(k)$:

$$\frac{\partial W}{\partial k} = 2\beta\alpha*(1-\alpha*) = 2\beta\alpha_a^*\alpha_b^* = \frac{\beta}{2} - \frac{\beta\theta_a^2}{2(t-2\beta(1-k))^2}$$

In cases where the welfare maximizing network is characterized by market sharing, the following condition is fulfilled: $t - 2\beta(1-k) > |\theta_a|$. Both the numerator and denominator are then positive and in such cases welfare is everywhere increasing in interconnect quality. The welfare maximizing interconnect quality is found by solving: $k* = \arg\max(W(k) - I(k))$. The first order condition is accordingly:

$$\frac{\partial W}{\partial k} - I'(k) = 0 \Leftrightarrow \frac{\beta}{2} - \frac{\beta\theta_a^2}{2(t-2\beta(1-k))^2} = I'(k)$$

The input joint venture will accordingly over-invest in interconnect quality when:

$$\beta - \frac{\beta\theta_i^2}{9(t-\beta(1-k))^2} > \frac{\beta}{2} - \frac{\beta\theta_a^2}{2(t-2\beta(1-k))^2} \Leftrightarrow$$

$$\frac{\beta}{2} + \beta\theta_i^2\left(\frac{1}{2(t-2\beta(1-k))^2} - \frac{1}{9(t-\beta(1-k))^2}\right) > 0$$

A sufficient condition is then that the large bracket is positive. This is the case since:

$$\begin{aligned}
&2(t-2\beta(1-k))^2 < 9(t-\beta(1-k))^2\\
&0 < 7t^2 - 10\beta t(1-k) + \beta^2(1-k)^2\\
&\quad = 7t^2 - 10\beta t(1-k) - 8\beta^2(1-k)^2 + 9\beta^2(1-k)^2\\
&0 < (t-2\beta(1-k))\underbrace{(7t+4\beta(1-k))}_{+} + \underbrace{9\beta^2(1-k)^2}_{+}
\end{aligned}$$

It is only socially optimal to set up a direct link between the two networks if both networks have a positive market share, this is the case when $(t-2\beta(1-k)) > |\theta_a|$. Thus the first bracket has to be positive. An input joint venture will accordingly over invest in interconnect quality under product differentiation as well as in the absence of vertical differentiation.

References

Armstrong, M., 1998. Network interconnection in telecommunications. The Economic Journal 108, 545–564.

Bailey, J., McKnight, L., 1997. Scalable internet interconnection agreements and integrated services. In: Kahin, B., Kellser, J.H. (Eds.), Coordinating the Internet. The MIT Press, Cambridge, Massachusetts.

Baake, P., Wichmann, T., 1998. On the economics of internet peering Netnomics 1, 89–105.

Cave, M., Crowther, P., 1999. Call Origination and Termination Charges for Accessing the Internet International Journal of Communications Law and Policy Issue 4, Winter 1999/2000.

Cawley, R.A., 1997. Interconnection, pricing and settlements: Some healthy josting in the growth of internet. In: Kahin, B., Kellser, J.H. (Eds.), Coordinating the Internet. The MIT Press, Cambridge, Massachusetts.

Chinoy, B., Solo, T.J., 1997. Internet exchanges: Policy-driven evolution. In: Kahin, B., Kellser, J.H. (Eds.), Coordinating the Internet. The MIT Press, Cambridge, Massachusetts.

Choi, S.Y., Stahl, D.O., Whinston, A.B., 1997. The Economics of Electronic Commerce. Macmillan Technical Publishing, USA.

Crémer, J., Rey, P., Tirole, J., 2000. Connectivity in the Commercial Internet. Journal of Industrial Economics XLVIII, 433–472.

DangNguyen, G., Penard, T., 1999. Interconnection between ISPs, Capacity Constraints and Vertical Differentiation, Paper presented at the 2nd Berlin Internet Workshop 28–29 May 1999.

Ergas, H., 2000. Internet Peering: A Case Study of the ACCC's Use of its Powers Under Part XIB of the Trade Practices Act, 1974. Paper presented at the 3rd Berlin Internet Workshop 26–27 May 2000, available at www.berlecon.de/iew3

Gareiss, R., 1999. The Old Boys' Network. Data Communications, October 1999

Gibbons, R., Mason, R., Steinberg, R., 2000. Internet Service Classes under Competition, forthcoming IEEE Journal of Selected Areas in Communications, available at www.soton.ac.uk/~ram2

Katz, M., Shapiro, C., 1985. Network externalities, competition and compatibility. American Economic Review 75, 424–440.

Kende, M., 2000. The Digital Handshake: Connecting Internet Backbones. OPP Working Paper No. 32, FCC.

Laffont, J., Tirole, J., 2000. Competition in Telecommunications. The MIT Press, Cambridge, Massachusetts.

Laffont, J., Rey, P., Tirole, J., 1998a. Network competition I: Overview and non-discriminatory pricing. Rand Journal of Economics 29 (1), 1–37.

Laffont, J., Rey, P., Tirole, J., 1998b. Network competition II: Price discrimination 29 (1), 38–56.

Little, I., Ralph, E., Warran, T., Wright, J., 2000. Welfare Analysis of International and Domestic Markets for Internet Access. Paper presented at the 3rd Berlin Internet Workshop 26–27 May 2000, available at www.berlecon.de/iew3

Mason, R., 2000. Simple Competitive Internet Pricing. European Economic Review 44 (4–6), 1045–1056.

Mason, R., 1999. Compatibility between Differentiated firms with Network Effects, Discussion Paper no. 9909, University of Southampton, available at www.soton.ac.uk/~ram2

Mueller, M., Hui, J.Y., Cheng, C., 1997. The Hong Kong exchange: The economics, evolution and connectivity of Asian internet infrastructure. In: Kahin, B., Kellser, J.H. (Eds.), Coordinating the Internet. The MIT Press, Cambridge, Massachusetts.

Rohlfs, J., 1974. A theory of independent demand for communications service. Bell Journal of Economics 5, 16–37.

Shapiro, C., Varian, H., 1998. Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press, Boston, Massachusetts.

Section 5

Demand-side Spillovers and Semi-collusion in the Mobile Communications Markets

Journal of Industry, Competition and Trade, 2:3, 259–278, 2002

Demand-side Spillovers and Semi-collusion in the Mobile Communications Market

ØYSTEIN FOROS* oystein.foros@nhh.no
Norwegian School of Economics and Business Administration (NHH), Depatment of Economics, Hellev. 30, N-5045 Bergen, Norway

BJØRN HANSEN bjorn.hansen@telenor.com
Telenor Research & Development, N-1331 Fornebu, Norway

JAN YNGVE SAND† jan.sand@nfh.uit.no
University of Tromso, Department of Economics and Management, NFH, N-9037 Tromso, Norway

Abstract. We analyze roaming policy in the market for mobile telecommunications. Firms undertake quality improving investments in network infrastructure in order to increase geographical coverage, capacity in a given area, or functionality. Prior to investments, roaming policy is determined. We show that under collusion at the investment stage, firms' and a benevolent welfare maximizing regulator's interests coincide, and no regulatory intervention is needed. When investments are undertaken non-cooperatively, firms' and the regulator's interests do not coincide. Contrary to what seems to be the regulator's concern, firms would decide on a higher roaming quality than the regulator. The effects of allowing a virtual operator to enter are also examined. Furthermore, we discuss some implications for competition policy with regard to network infrastructure investment.

Keywords: mobile communications, roaming, competition, virtual operators

We analyze competing mobile telephony providers' incentives to invest in, and share infrastructure. Furthermore, we analyze whether the regulator should intervene in the firms' sharing agreements, and whether the regulator should allow firms to coordinate their investments. The infrastructure investment we have in mind is an upgrade of the mobile networks from second generation (2G) to third generation (3G) systems. The particular class of agreements we analyze is called roaming agreements.

The main improvement of 3G networks (*e.g.* UMTS) compared to the current 2G mobile networks (*e.g.* GSM) is to increase the speed of communication in the access network and thereby give access to new services and new functionality for existing services. Investments will therefore increase consumers' willingness to pay for mobile access. The basic mechanism driving our results is that in the presence of roaming

* Corresponding address: Norwegian School of Economics and Business Administration, Department of Economics, Hellev. 30, N-5045 Bergen, Norway.

† This work has been carried out as part of a research project at Telenor R&D—DN 0401. The authors are grateful to the editor and an anonymous referee for comments. Furthermore, we thank Kåre P. Hagen, Derek Clark, Lars Mathiesen, Lars Sørgard and Tim Brennan, participants at the 19th Annual Conference on the ''Competitive challenge in network industries'', May 2000, Lake George, New York, EUNIP 2000, December 2000, Tilburg, Netherlands, Helsinki Conference on Anti-trust Issues in Network Industries 2001, and EARIE 2001, Dublin, Ireland, for comments and discussions.

agreements, investments carried out by one firm increase the value of the services provided by other firms.

An analysis of the consequences of coordinating investments and infrastructure sharing seems to be more relevant in 3G as compared to 2G networks. First, while the providers of 2G networks (GSM) typically made their investment non-cooperatively, we now see that several providers of UMTS coordinate their investments in infrastructure (*e.g.* in Sweden and Germany). There has been a heated debate whether the regulator should allow the firms to cooperate at the investment stage. Second, the benefit from sharing agreements through roaming seems to be higher in 3G as compared to 2G networks. In current 2G networks the consumers have access to a given capacity (9.6 kbit/s), while in 3G the available capacity for data transmission can be allocated in a more dynamic way. If there are free resources in the network a consumer may be given a capacity of up to 2 megabit/s, but as the number of users in a given area at a given time increases, each user will have less capacity available. This may increase the value of infrastructure sharing agreements. There are potential gains from sharing network capacity when the load in one network is not perfectly correlated with the load in other networks. Let us illustrate this by a simple example. Let us assume that at a given time, operator A has no free capacity in its network whereas operator B has idle capacity. Suppose now that a subscriber of A tries to download capacity intensive content and needs 2 megabit/s. If there is no capacity sharing agreement between A and B, the customer will not be able to access such services at that time. If a sharing agreement is established between the operators, the service will be available to the consumer. In this situation, it is obvious that an investment in capacity by B will increase the willingness to pay for subscriptions from firm A.

We analyze a stylized multi-stage model where the firms first agree on roaming quality, second choose their investment non-cooperatively or cooperatively, and finally compete à la Cournot. We investigate whether it is welfare improving to let the firms semi-collude by choosing their investment cooperatively before they compete in the downstream market. Alternatively, we may have semi-collusion where the firms compete at the investment stage and collude in the retail market such as in Brod and Shiwakumar (1999), Fershtman and Gandal (1994), and Steen and Sørgard (1999). The latter form of semi-collusion seems realistic since the firms typically will collude on the most observable variable. This will usually be the price in the retail market. Our motivation for a different type of semi-collusion is that we observe collusion at the investment stage in the market we consider, and we want to investigate the effects of allowing such collusion.[1] Therefore, we do not consider collusion in the final product market.[2]

There are two symmetric facility-based firms in our basic model. If the investments are set non-cooperatively we show that voluntary roaming leads firms to agree on a too high roaming quality compared to the social optimum. Moreover, investments are strategic

1 In addition, general competition rules (usually) prohibit collusion in the final product market whereas investment cooperation may be allowed.

2 See Busse (2000) and Parker and Röller (1997) for analysis of tacit collusion in the mobile market.

complements and firms will accordingly invest less with voluntary than with mandatory roaming. In contrast, if the investments are set cooperatively, the firms' choices on the roaming quality coincide with the regulator's interests. We show that it is welfare improving to allow the firms to semi-collude in the way described above.

In an extension of our basic model we introduce a non-facility-based firm, or a virtual operator, in addition to the two facility-based firms. Whether a virtual operator should be allowed to enter the market, and to which extent the presence of such an operator will affect the incentives to invest in infrastructure, has been heavily debated in the industry and amongst regulators. This debate started when the Scandinavian virtual mobile operator Sense Communication attempted to get access to the facility-based mobile operators' networks. The facility-based firms were reluctant to grant Sense Communication access.[3] A much more friendly reception was given to Virgin Mobile in the United Kingdom.

The roaming quality between the entrant and the two facility-based firms in our model is assumed to be weakly lower than the roaming quality between the two facility-based firms. We show that when the investments are set non-cooperatively between the facility-based firms an increase in the roaming quality of the incumbents may increase the investments. This is in contrast to the results in the basic model.

We analyze a type of semi-collusion where the firms may collude at the investment stage, but always compete in the retail market. Our model is an extension of the models of d'Aspremont and Jacquemin (1988) and Kamien et al. (1992) which consider (strategic) R&D investment. The investments in infrastructure give rise to spillover effects, through the roaming agreements, similar to those considered in models of strategic R&D investments. In the majority of these models, the externality is exogenous. In our model, we focus on the situation where the level of the externality is endogenously determined.[4] In contrast to the majority of the R&D literature we introduce asymmetry between the firms, with both investing firms and firms that do not invest in infrastructure.[5]

We also make some other key assumptions in our model. First, for the sake of simplicity we ignore the issue of interconnection (agreements that give access to rivals' customer bases) and focus solely on roaming. We give roaming a wider interpretation than pure geographical coverage. Roaming agreements extend availability, such that (i) subscribers can make and receive calls via the infrastructure coverage of a rival operator,

3 Sense wished to issue their own SIM cards, but the Scandinavian facility-based operators refused this. Sense filed a complaint to all national regulators, but only the Norwegian regulator supported it. Telenor's appeal to the Norwegian regulator was still pending when Sense filed for bankruptcy in March 1999 (Matthews, 2000). Now, Sense Communication, along with several other virtual operators, has an agreement with Telenor to resell airtime.

4 Katsoulacos and Ulph (1998a,b) introduce endogenous levels of spillovers between firms. Contrary to their models we assume that the investments undertaken by firms result in product innovation with probability one. Furthermore, our focus is on a context where firms (or a regulator), in the terminology of Katsoulacos and Ulph, choose the degree of information sharing and not research design. We are accordingly examining investments with firms operating in the same industry, but pursue complementary research.

5 See *e.g.* De Bondt (1997) for a survey of the R&D literature of strategic investments.

(ii) when there is congestion a customer may take advantage of the infrastructure of the rival, and (iii) give access to new functionality/services in the rivals' networks.

Second, to simplify we make the assumption that there are no side payments between firms engaged in roaming. If firms have the ability to write complete contracts in all dimensions of infrastructure sharing (*i.e.* roaming), all external effects from the investment can in principle be internalized through the price mechanism. Then, the problem of spillovers through roaming analyzed in this paper may vanish. Since sharing agreements in the next generation systems aim at ensuring a more dynamic capacity allocation, it is rarely possible to write complete contracts in all dimensions. This implies that even if a price mechanism for roaming is implemented, it will not be able to internalize all external effects. This is similar to what we see in the Internet, where infrastructure sharing of backbones is common (Crémer et al., 2000). Note that regulation may also constrain the firms' ability to internalize external effects through pricing. In particular, this will be important in the interaction between the facility-based firms and the non-facility based firm (virtual operators).

Third, we make the simplifying assumption that consumers only pay for subscription, not for usage. This is evidently a restricting assumption since mobile providers typically employ various types of nonlinear pricing, but it is far from evident what the alternative is, and in particular, whether mobile providers will choose to price discriminate between calls originated off-net and calls originated on-net.[6] The focus in our model, however, is on how availability in various dimensions (capacity, speed *etc.*) affects the choice of supplier. Our focus is on demand for subscriptions, and then it is sufficient to consider pricing as a fixed per period fee.

Fourth, we assume Cournot competition in the retail market. We interpret the quantity firms dump in the retail market as the number of subscriptions they sell. A justification for assuming Cournot competition is that there are both physical and technological limits to capacity, due to the fact that the amount of radio spectrum available is scarce. Furthermore, firms must choose a capacity level (which is built or rented) in both the backbone network and the access network (number of base stations) prior to the competition in the retail market.[7] This will be more important with 3G systems where the capacity needed increases. However, as shown by Kreps and Scheinkman (1983), strong

6 In addition to the pricing issues, we will have to model the cost structure for calls originating and terminating on the same net by a subscriber of that network if a call volume dimension is introduced. Under such a generalization of our model it would also be natural to relax assumption 2 and introduce a volume price on roaming (as well as a volume price on interconnection). A proper modeling of all these pricing and cost components will lead to a very complex model. By disregarding both the revenue and the cost side of traffic we also avoid the so-called ''bill and keep fallacy''.

7 The basic structure in a mobile network is that coverage in a given area is achieved through a number of base stations covering given areas (cells). Hence, a mobile network consists of a network of such cells. The spectrum band allocated for mobile use limits the total bandwidth a cell can handle at a given point of time. Thus total capacity measured is limited and one bandwidth hungry user occupying 2 Mbit/s is crowding out approximately 200 ordinary voice calls. If there are capacity problems, it is possible to increase capacity through what is called cell splitting. Cell splitting implies that a given area is served with a higher number of smaller cells.

assumptions are required to ensure that a capacity constrained price game result in identical results as a Cournot game. Nevertheless, this seems more appropriate than assuming a Bertrand game without capacity limits.

The rest of the paper is organized as follows: In Section 1, the model with only facility-based firms is presented and analyzed. In Section 2, we provide an extension to the basic model by introducing a virtual operator. In Section 3, some concluding remarks are made.

1. The model

In our basic model we will look into a duopoly case where we assume the following three-stage game:

Stage 1: Choice of roaming quality (decision taken by either the firms or the regulator)
Stage 2: The firms determined infrastructure investments either non-cooperatively or cooperatively
Stage 3: Cournot competition

There are four different variants of the game depending on the stage 1 and 2 strategies (see Figure 1).

The choice of whether firms cooperate when determining their investment levels will depend on whether such cooperation is approved by the competition authorities. Stages 2 and 3 in our model is fairly similar to the structure in Kamien et al. (1992) and d'Aspremont and Jacquemin (1988). The generic feature of the investment is that it leads to product innovation, which increases the quality of the services provided.

One interpretation of the timing in our model is that roaming policy may be part of the licenses to the operators in the case where the degree of roaming is mandatory, and therefore chosen prior to investments taking place. When roaming is voluntary, we assume that firms can commit to a policy on roaming prior to undertaking the investments. Indeed, the timing of the roaming quality decision relative to infrastructure investments can obviously be different. To be more specific, the infrastructure investment may be decided prior to a decision on roaming quality. Such timing may involve

		Stage 2: Investments	
		Competitive	Collusive
Stage 1: Roaming	Voluntary	Game 1	Game 3
	Mandatory	Game 2	Game 4

Figure 1. The four variants of the game.

problems with investment hold-ups, but this will not be our main focus. In our choice of timing we implicitly assume that the firms/the regulator can credibly commit to a given policy on roaming. As far as the regulator is concerned, the issued licenses may serve as a commitment device, whereas the commitment problem under voluntary roaming is solved, *e.g.* if a given roaming policy is embedded in the network design (*e.g.* type of interfaces).[8]

The demand side. When firm i invests in infrastructure it impacts on the quality of the services its own customers are offered, but there may also be an impact on the quality of the services offered by the rival firm j, and vice versa. Given a roaming policy β and investment decisions x_i, we can now write the total quality offered to consumers by firm i:

$$a_i = a + x_i + \beta x_j \tag{1}$$

where x_i is the network investment undertaken by firm i, and x_j indicates the investment by the rival. We assume that $\beta \in [0, 1]$ is a parameter indicating the degree of roaming. This parameter measures the externality effect from sharing infrastructure. If $\beta = 1$, there is an agreement on full roaming, while $\beta = 0$ implies minimum roaming quality.

The inverse demand function faced by firm i is given by:

$$p_i = a_i - q_i - q_j$$

The price, p_i, is the subscription fee (*i.e.* a monthly fee). The externality introduced above is such that when firm i invests in infrastructure, the marginal willingness to pay for the final products produced by both firms is increasing.

The supply side. We assume a linear cost function in the final stage for firm i given by $C_i = cq_i$. The cost c is the direct cost associated with access connection of one user. We assume that firms face quadratic (network infrastructure) investment costs, given by $TC_i(x_i) = \varphi x_i^2/2$, where $\varphi > 4/3$. We will later demonstrate that the restriction on φ ensures a unique and stable equilibrium. Overall profit for firm i is then:

$$\pi_i = (p_i - c)q_i - \frac{\varphi x_i^2}{2} \tag{2}$$

8 Poyago-Theotoky (1999) considers a model of R&D where the degree of spillover is endogenous. In her model, the timing of the game is different from ours, in that the R&D investment decision (which is equivalent to our infrastructure investment decision) is made prior to the decision on how much information to share with competitors. In addition, she allows firms to choose different levels of spillover. In our model, the degree of spillover (interpreted as roaming quality) is reciprocal, in that the degree of spillover is identical in both directions. When firms choose R&D cooperatively they choose to fully disclose their findings, whereas when there is competition in R&D firms choose minimal disclosure. The latter result is very different from what we find in our model.

Welfare. We assume that the regulator maximizes welfare given by the sum of producer and consumers' surplus:

$$W = CS + \sum_{i=1}^{2} \pi_i \tag{3}$$

Since firms are symmetric and the inverse demand functions are linear with identical slopes, we can write consumers' surplus as $CS = 2q^2$, where q is the symmetric production level of each firm.

Cournot competition (stage 3). At stage 3, firm i maximizes the profit function: $\pi_i = (p_i - c_i)q_i$. Combining the first order conditions for the two firms we obtain the equilibrium quantities:

$$q_i^* = \frac{a - c + x_i(2-\beta) + x_j(2\beta - 1)}{3}$$

Note that in a symmetric equilibrium $(a_i = a_j)$, the equilibrium quantity is given by $q^* = (a_i - c)/3$.[9] This quantity is monotonously increasing in $a_i = a_j$, and this implies that consumers' surplus is monotonously increasing in a_i. Firm i obtains stage 3 equilibrium profits given by $\pi_i = (q_i^*)^2$.

1.1. Infrastructure investment (stage 2)

When firms invest in infrastructure at stage 2 of the game, they take into account the effect such investments has on the stage 3 equilibrium.

1.1.1. Non-cooperative solution

At stage 2, the firms maximize the profit function (2), subject to Cournot equilibrium quantities at stage 3, which implies that the (symmetric) equilibrium investment is given by:

$$x_{nc}^* = \frac{(4 - 2\beta)(a - c)}{9\varphi - 2(2-\beta)(1+\beta)} \tag{4}$$

In equilibrium, firms' profits are non-negative for all permissible parameter values, and firms will participate in the game. Furthermore, the symmetric equilibrium is the unique equilibrium.[10] Our result is analogous to d'Aspremont and Jacquemin (1988) and Kamien et al. (1992), and is summarized in the following lemma:

9 As it turns out, the unique equilibrium in investment is indeed the symmetric equilibrium.

10 The second-order condition requires that $2(2-\beta)^2 - 9\varphi < 0$. In order to have a stable equilibrium the slope of the reaction function has to have an absolute value below unity. It is straightforward to demonstrate that this condition is fulfilled for $\varphi > 4/3$. Hence, the second order condition is also fulfilled when $\varphi > 4/3$. Finally, it can be shown that the symmetric equilibrium is indeed unique.

Lemma 1: If $\beta > 1/2$, then x_i and x_j are strategic complements; that is, $(\partial x_i/\partial x_j)^{nc} > 0$. Reversing the inequality makes x_i and x_j strategic substitutes.

When one firm invests in its infrastructure, the equilibrium quantity of the other firm may or may not increase as a result of the spillover from the roaming agreement. As pointed out by Kamien et al. (1992), the spillover externality (or investment externality) is positive if and only if $\beta > 1/2$. In other words, the spillover externality is positive only when x_i and x_j are strategic complements.

1.1.2. Cooperative solution

In the cooperative case (collusion) the two firms coordinate their infrastructure investments at stage 2, and compete à la Cournot at stage 3 in the same way as above. When determining the profit-maximizing choice of investments at stage 2, firm i maximizes the joint profit of the two firms (*i.e.* the industry profits):

$$\max_{x_i} \Pi = (q_i)^2 + (q_j)^2 - \frac{\varphi}{2}(x_i^2 + x_j^2) \quad \text{for} \quad i \neq j \tag{5}$$

The following first-order condition yields the equilibrium investment for a given firm under collusion:[11]

$$x_c^* = \frac{2(a-c)(1+\beta)}{9\varphi - 2(1+\beta)^2} \tag{6}$$

The cooperative solution yields lower infrastructure investment than the non-cooperative, $x_c^* < x_{nc}^*$, if and only if $\beta < 1/2$. This is equivalent to the results obtained by d'Aspremont and Jacquemin (1988) and Kamien et al. (1992).

Observe that for $\beta < 1/2$, equilibrium quantity in the final stage of the game is lower under collusion relative to the non-cooperative case, which implies that consumers' surplus is lower under collusion relative to non-cooperation if the roaming quality. If $\beta > 1/2$, consumers' surplus is higher if there is collusion at the investment stage. In addition, firms' profits under collusion are always at least as large as under non-cooperation.[12] A welfare-maximizing regulator would therefore choose to allow collusion at stage 2, and such collusion would be in the firms' interests provided that the roaming quality is sufficiently high (*i.e.* $\beta > 1/2$). In the next section we examine which level of roaming quality a welfare-maximizing regulator would choose.

11 In the same way as in game 1, the second-order condition, $2(1+\beta)^2 - 9\varphi < 0$, is satisfied with our parameter restrictions.

12 For $\beta = 1/2$, firms are indifferent between collusion and non-cooperation at stage 2.

1.2. Roaming quality (stage 1)

In this section, we extend the model by considering the two cases where: (1) Firms decide roaming quality (*voluntary roaming*) and (2) the regulator decides roaming quality (*mandatory roaming*). These two cases combined with the two ways of determining investments at stage 2 yields four different games (see Figure 1). The analysis in this section can be seen as extending the basic model with exogenous R&D spillovers to enable us to analyze endogenous spillovers.

1.2.1. Game 1, voluntary roaming when investments are determined non-cooperatively

Recall that equilibrium infrastructure investment (under non-cooperation) is given by Equation (4). Direct differentiation of (4) yields the following result:

Lemma 2: When firms determine the investments non-cooperatively, the infrastructure investment decreases as the roaming quality increases since $\partial x^*_{nc}/\partial\beta < 0$.

The intuition behind Lemma 2 is as follows: We are considering the equilibrium with reciprocal spillover levels, implying that both firms' final products increase in value to consumers by the same proportion. Hence, there is no product differentiation gain from investments for any of the firms. When firms do not cooperate at the investment stage, they cannot internalize the effect of the investment on the other firm's profit. An increase in the degree of roaming quality therefore affects the marginal revenue from investing in infrastructure adversely, since the investing firm is unable to capture the effect on the rival's profit. When examining the stage 3 equilibrium we observe that equilibrium quantities increase in both the degree of roaming quality and the investment level. Since quantities are strategic substitutes, firm j will increase its production if firm i reduces his. Each firm will then have to be more cautious and ration its production of the final product more than is the case if they collude in the investment stage. To achieve a substantial enough reduction, any given firm will have to limit its production even more. Since q_i is increasing in β and $\left(q_i + q_j\right)$ increases when x_i increases, the investing firms can ration the final product market by reducing the level of investments when the degree of roaming quality increases.

The competitive equilibrium infrastructure investment is given by Equation (4); $x^*_{nc} = x_{nc}(\beta)$. From Lemma 2 we know that $x'_{nc}(\beta) < 0$. By symmetry, the equilibrium profit is accordingly:

$$\pi = \left(\frac{a + (1+\beta)x_{nc}(\beta) - c}{3}\right)^2 - \varphi\frac{(x_{nc}(\beta))^2}{2} \tag{7}$$

Differentiating equilibrium profit with respect to β, reveals the roaming quality preferred by the firms.

$$\frac{2}{9}(a - c + (1+\beta)x_{nc}(\beta))(x_{nc}(\beta) + (1+\beta)x'_{nc}(\beta)) - \varphi x'_{nc}(\beta) = 0 \tag{8}$$

It can be shown that the optimal β for the firms is independent of $(a-c)$. The solution to Equation (8) will be a function of the convexity of the investment cost function, *i.e.* of the value for φ. The profit maximizing choice with respect to β is increasing and concave in φ and strictly larger than 9/10. As an example, when $\varphi = 3/2$ the expression given by Equation (8) is concave over the interval where it is defined, and the first order condition is satisfied for $\beta = 0.941$ The firms will accordingly agree on $\beta = 0.941$ as the preferred level of roaming, if roaming is voluntary.

Even if firms can agree on the level of roaming that maximizes profit, there may be a commitment problem for the firms. After firms have chosen their level of expenditure on network infrastructure (at stage 2), both firms have incentives to renegotiate the roaming quality between stages 2 and 3. The reason for this is that for a given level of x_i, the (stage 3) Cournot-equilibrium profit of both firms is strictly increasing in β. If firms cannot commit to the roaming quality chosen at stage 1 of the game, we may experience hold-up problems in network infrastructure investments.[13] As indicated earlier, this commitment problem is solved if roaming policy is embedded in the network design, such that the firms cannot change roaming policy after the investments have been made. Furthermore, in Section 2 we demonstrate briefly that the introduction of a virtual operator can eliminate the commitment problem for the firms.

1.2.2. Game 2, mandatory roaming when investments are determined non-cooperatively

In this section we investigate the roaming quality choice of a welfare-maximizing regulator, which can either be considered as a benchmark case or as the chosen roaming quality under mandatory roaming. We assume that the regulator maximizes the objective function given by Equation (3). Producer surplus is calculated above (Equation (7)). Consider now the roaming quality preferred by consumers. We know that the equilibrium is symmetric in the sense that the two firms invest in the same level of infrastructure and that they offer the same quantity at stage 3 of the game. Let x^* and q^* denote the profit-maximizing choices of investment and quantity, respectively. In the equilibrium, we have the following equilibrium price:

$$p = a + (1+\beta)x^* - 2q^*$$

Inserting for the non-cooperative equilibrium investment given by Equation (4), and $q^* = ((a-c) + (1+\beta)x_{nc}^*)/3$ the consumers' surplus becomes:

$$CS_{nc} = \frac{2[(a-c) + (1+\beta)x_{nc}(\beta)]^2}{9} \tag{9}$$

13 Since the firms' profit is increasing in roaming quality for a given level of investment, firms have incentives to increase β to its maximum after the investments are sunk. The hold-up problem arises because we know that when investments are made non-cooperatively, the marginal revenue of the investments is reduced when β increases. Thus, if firms know that, ultimately, roaming quality is chosen to give perfect roaming, they will hold back on investments.

The first-order condition for Equation (9) with respect to β is satisfied for roaming quality $\beta = 0.5$,[14] which is the roaming quality that maximizes consumers' surplus.[15]

The roaming quality that maximizes consumers' surplus can be compared to the roaming quality that maximizes the profit of the firms (Equation (8)). In Section 1.2.1 we demonstrate that the roaming quality that maximizes producer surplus is strictly larger than 9/10. It is then evident that the firms prefer a roaming quality that exceeds the roaming quality preferred by consumers. Since welfare is the sum of consumer and producer surplus, the welfare maximizing roaming quality (game 2) is below the level preferred by the firms (game 1). This result is summarized in Proposition 1:

Proposition 1: Assume that investments are undertaken non-cooperatively. Voluntary roaming induces firms to choose a higher level of roaming quality relative to the social optimum. Consequently, in a voluntary roaming regime firms invest less in network infrastructure than in a mandatory regime.

The intuition behind this result is as follows: Consumers will be better of with high levels of infrastructure investments, for a given quantity of the final product. The investing firms cannot capture all of the benefits from the investments, since the presence of a roaming agreement implies that there are external effects from the investments (which is internalized by a welfare-maximizing regulator). In addition, firms face convex costs of investing. This implies that firms will choose an investment level that is lower than the level that maximizes consumers' surplus. From Lemma 2 we know that incentives to invest in infrastructure are worse when the roaming quality is high, since each of the investing firms will attempt to be a free rider on the other firm's investments. This leads to a lower equilibrium investment level. This implies that when roaming is voluntary firms choose a high level of roaming quality to reduce the investment levels. When roaming is mandatory, a lower level of roaming quality is chosen to induce higher levels of investment in network infrastructure.

If firms are not allowed to collude at the investment stage, regulatory intervention may be needed. In this case, firms would set a higher roaming quality than the socially optimal level, or equivalently, firms would agree on more compatibility than what would be beneficial to consumers and society as a whole. As a result, the output of the final product in stage 3 of the game is also restricted, with an increase in equilibrium price. For consumers, an increase in roaming quality has two potentially opposing effects. Increased roaming quality implies (under non-cooperation) that investment is reduced. This implies that the size of the market is reduced (the inverse demand function is shifted inwards). For a given quantity, this results in a reduction in the price charged to consumers and has a positive impact on consumers' welfare. A reduction in the level of infrastructure investment results in a reduction in quantities sold in the last stage of the game, which has

14 The solution that $\beta = 0.5$ maximizes consumers' surplus is also independent of the convexity of the investment cost function, *i.e.* of the value for φ.

15 The second-order condition for CS is satisfied for all permissible values of β.

a negative impact on consumers' welfare. The overall effect on consumers' welfare of an increase in the quality of roaming is positive for low levels of roaming quality, and negative for high levels.

1.2.3. Games 3 and 4, voluntary and mandatory roaming when investments are determined cooperatively

Recall that the equilibrium investment as a function of roaming quality is given by Equation (6). Direct differentiation yields the following result:

Lemma 3: When firms collude at the investment stage, the infrastructure investment increases as the roaming quality increases since $\partial x_c^*/\partial\beta > 0$.

Contrary to the case of non-cooperative investment, the incentives to invest are in fact higher when the roaming quality is high. In this case, firms achieve a better coordination due to the fact that the effect on the other firm's profits is internalized and all benefits from the investments are credited to the investing firm. This changes investment incentives qualitatively. There is no longer the problem that each of the investing firms will have incentives to free ride on the other firm's investments. If firms determine a high (low) level of roaming quality, each firm's investment presents a large (minor) positive external effect the firms jointly can internalize. By inserting for the collusive equilibrium investment, x_c^* (given by Equation (6)), and third stage equilibrium quantities, we obtain consumers' surplus (under collusion), CS_c:

$$CS_c = 2\left[\frac{a-c}{3} + \frac{2(a-c)(1+\beta)^2}{3(9\varphi - 2(1+\beta)^2)}\right]^2 \tag{10}$$

When firms decide on infrastructure investments collusively, consumers' surplus given by Equation (10) is increasing and convex in roaming quality over the relevant interval for β, which implies that the optimal β for consumers is equal to unity (or maximal roaming quality). Quantity is increasing in roaming quality both directly and through the effect roaming quality has on investment incentives. Consumers' surplus is increasing in quantity, such that the higher the roaming quality the higher is the level of consumers' surplus.

Firms maximize their equilibrium profit under collusion with respect to β, which results in:

$$\frac{\partial\pi}{\partial\beta} = 2q_c^*\left[\frac{x_c^*}{3} + \left(\frac{1+\beta}{3}\right)\frac{\partial x_c^*}{\partial\beta}\right] - \varphi x_c^*\frac{\partial x_c^*}{\partial\beta} \tag{11}$$

where q_c^* is the stage 3 equilibrium quantity for each firm if firms collude at the investment stage. In the collusion case, the profit function is increasing and convex in β over the interval $[0, 1]$. This implies that the optimal roaming quality for firms corresponds to the maximal roaming quality. We summarize our findings in Proposition 2:

Proposition 2: Assume that firms collude at stage 2. The (unregulated) profit maximizing choice of roaming quality is identical to the socially optimal roaming quality. Consequently, the level of investments in infrastructure is identical in both the voluntary and mandatory roaming regimes.

Since firms' and consumers' interests coincide, there is no reason for governmental intervention and there is no need for considering games 3 and 4 separately. As in the non-cooperative case, consumers benefit from high investments in infrastructure for a given production level. To achieve high levels of investments, consumers must now choose a high level of roaming quality. The main reason for the difference in our result is that firms, when allowed to coordinate their investments, are able to capture all benefits from the investments. Because of this, firms' investment incentives are changed and they now seek high roaming quality to induce high levels of investments, whereas in the non-cooperative case they seek high roaming quality to induce low levels of investments.

1.3. Collusive and competitive investments compared

At stage 2 of the game the level of investment in the network infrastructure is determined. As already stated, firms can either compete (games 1 and 2) or they can collude when determining the investment level (games 3 and 4). We may think of the decision to allow firms to collude or not as a decision taken by the competition authorities prior to commencement of the 3-stage game we analyze above.

Assuming that roaming is voluntary we can compare equilibrium under collusive and competitive investments, respectively (*i.e.* we compare games 1 and 3), to determine whether the investing firms should be allowed to collude or not at the investment stage. Under voluntary roaming the firms set the β to maximize profits. Firms are evidently better off under collusion as compared to competitive investments.[16] Furthermore, it follows from the calculations above that equilibrium consumers' surplus is higher under collusion (games 3 and 4) as compared to equilibrium consumers' surplus under competitive investments and mandatory roaming (game 1).[17] The intuition behind this result is as follows: Consumers' surplus increases, both in roaming quality for given investments and in investment for a given roaming quality. Since we have demonstrated that roaming is at the highest possible level under collusive investments and that investments are higher under collusion provided that $\beta > 0.5$, consumers' surplus will indeed be higher under collusion.

16 When firms collude they can always mimic the outcome under competition. If they choose to deviate from this outcome it is because they are better off.

17 This result is derived by first observing that consumers' surplus (CS) is higher in game 2 as compared to game 1. Then, a sufficient condition for demonstrating that CS is higher in games 3 and 4 as compared to game 2 is to assume that the regulator in game 2 determine the roaming at the level maximizing CS ($\beta = 0.5$). Then we can compare CS under collusion, Equation (10), for $\beta = 1$ and compare it CS under competitive investments, Equation (9), for $\beta = 0.5$. Then we find that $CS_{\text{games 3 and 4}} \geq CS_{\text{game 2}} \geq CS_{\text{game 1}}$.

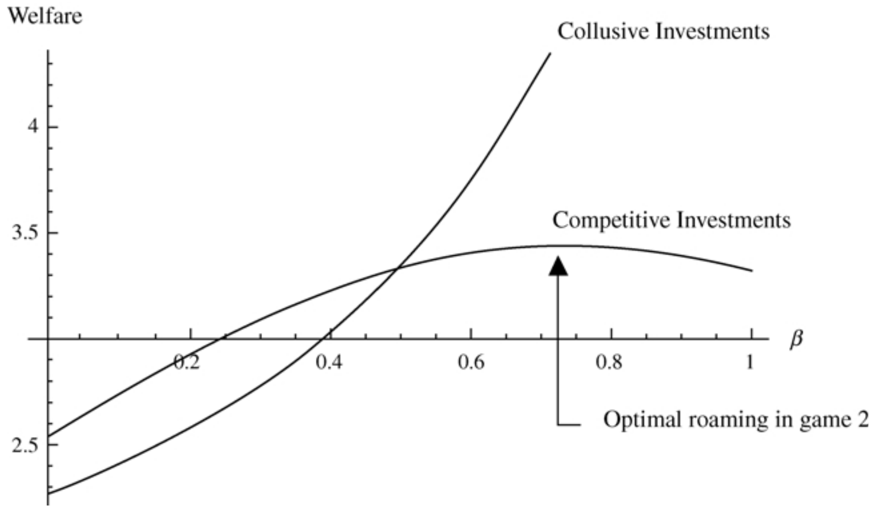

Figure 2. Welfare comparison of collusion versus non-cooperation.

In Figure 2 we illustrate the welfare effects of chosen roaming policy for a regulator under the two investment regimes.[18] It is evident from Figure 2 that social welfare is maximized when the regulator allows firms to collude at stage 2, provided that the quality of roaming is sufficiently high (*i.e.* for β higher than 0.5). We know from the analysis above that a welfare maximizing regulator indeed will set the quality of roaming equal to unity when firms collude at the investment stage. Figure 2 also illustrates the (perhaps) counter intuitive results that it is detrimental to welfare to choose roaming quality at too high a level under competitive investments.

We observe in our model that the competition authorities (or a regulator) can never do worse than allowing collusion at the investment stage, provided that the firms/regulator can commit to a sufficiently high roaming quality. The reason is that allowing collusion allows the positive external effects to be internalized.

2. Entry of a non-facility-based firm

In this section we analyze a setting with two different types of firms. One type is facility-based and invests in its own infrastructure. We assume that there are two facility-based incumbents as in the previous section. The other type is a virtual operator, who is an entrant without its own infrastructure. The timing of the game is the same as in the analysis above, but at stage 2 of the game it is only active firms are the facility-based firms. We assume that the facility-based firms have all the bargaining power in determining the quality of roaming for the virtual operator.

18 The parameter values are: $a = 3, c = 1$, and $\varphi = 3/2$.

The model is amended to incorporate the two types of firms. In the case where the entrant is a virtual operator we have:

$$a_i = a + x_i + \beta^f x_j$$
$$a_v = a + \beta^v (x_i + x_j)$$

where subscript $i, j = 1, 2, i \neq j$ represents the two incumbents, and subscript v represents the virtual operator.[19] The parameter β^f represents the degree of the roaming quality between the facility-based firms, while β^v is the roaming quality between a facility-based operator and a virtual operator. We restrict the analysis to the case where $0 \geq \beta^f \geq \beta^v$.[20] The parameters β^v and β^f can also be interpreted as the virtual operator's and facility-based operators' capabilities of transforming the inputs into a final product of high quality.

The profit function of a facility-based firm is given by Equation (2), while the profit of the virtual operator is $\pi_v = (p_v - c)q_v$, where $p_v = a_v - \sum_{i=1}^{2} q_i - q_v$. By combining the first-order conditions from stage 3 for the three firms we obtain the following equilibrium quantities:

$$q_i^* = \frac{[(a-c) + x_i(3 - \beta^f - \beta^v) + x_j(3\beta^f - \beta^v - 1)]}{4}$$
$$q_v^* = \frac{[(a-c) + (x_i + x_j)(3\beta^v - \beta^f - 1)]}{4}$$

2.1. *Incumbents set the investments non-cooperatively*

When the two facility-based firms set their investment non-cooperatively we find the (symmetric) equilibrium investment for each facility-based firm:[21]

$$x = \frac{(a-c)(3 - \beta^f - \beta^v)}{8\varphi - 2(3 - \beta^f - \beta^v)(1 + \beta^f - \beta^v)} \tag{12}$$

We can now examine how the equilibrium investment level x changes when β^f and β^v change:

Proposition 3: When introducing a virtual operator and when the investments are made non-cooperatively by the facility-based firms, an increase in the roaming quality between

19 Ceccagnoli (1999) gives a similar formulation with process innovation. In contrast to our model, he only focuses on the case where the R&D investment is set non-cooperatively.

20 This seems to be an appropriate assumption if the virtual operator is simply a reseller of airtime (*e.g.* Sense Communication in Scandinavia). However, if the virtual operator has a well known brand name and possesses detailed knowledge about certain segments of the market (*e.g.* Virgin in the United Kingdom), it may be reasonable to assume that investments benefit the virtual operator more than they benefit the incumbents.

21 The second order condition is fulfilled if $\varphi > 9/8$, which also ensures stability.

the investing firms may result in a higher equilibrium investment level; *i.e.* $\partial x/\partial\beta^f > 0$ if and only if $(\beta^f + \beta^v) < 2\sqrt{\varphi} - 3$. Furthermore, an increase in the roaming quality to the virtual operator reduces investment incentives, $\partial x/\partial\beta^v < 0$.

Proof:

$$\frac{\partial x}{\partial\beta^f} = \frac{-(a-c)[8\varphi - 2(3-\beta^f-\beta^v)^2]}{[8\varphi - 2(3-\beta^f-\beta^v)(1+\beta^f-\beta^v)]^2}$$

$$\frac{\partial x}{\partial\beta^v} = \frac{-(a-c)[8\varphi + 2(3-\beta^f-\beta^v)^2]}{[8\varphi - 2(3-\beta^f-\beta^v)(1+\beta^f-\beta^v)]^2} < 0$$

The first expression is positive if and only if $[8\varphi - 2(3-\beta^f-\beta^v)^2] < 0$. This implies that $\beta^f + \beta^v < -3 \pm 2\sqrt{\varphi}$. Since $\beta^f + \beta^v \in [0,2]$ the relevant root that satisfies the inequality is $(\beta^f + \beta^v) < 2\sqrt{\varphi} - 3$. ■

Recall the basic model with only two facility-based firms. In such a setting, the incentives to invest are lower the higher the degree of roaming quality if investments were undertaken non-cooperatively. For sufficiently high values of the roaming quality we obtain a similar result in the presence of a virtual operator. We observe from Proposition 3 that contrary to the basic model the incentives to invest may in fact be improved the higher the roaming quality is between the facility-based firms. This will be the case when $(\beta^f + \beta^v)$ is sufficiently small and φ is sufficiently large.

In stage 1 of the game, the level of roaming quality is chosen either by the investing firms or by the regulator. Our findings suggest that if roaming is voluntary and invest-ments are made non-cooperatively, the investing firms will choose to set the roaming quality between the investing firms as high as possible $(\beta^f = 1)$. The roaming quality between an investing firm and the virtual operator is set as low as possible $(\beta^v = 0)$.

It may, however, be the case that profit for the virtual operator is negative in this solution. A sufficient condition for ensuring non-negative profits for the virtual operator is that the convexity of the investment cost function is sufficiently large (*i.e.* φ being sufficiently large). It can be shown that the restriction $\varphi > 2$ ensures non-negative profits for the virtual operator under competitive investments. The intuition behind this result is that when $(3\beta^v - \beta^f - 1) < 0$ any investments in infrastructure undertaken by the facility-based firms will reduce the equilibrium output of the virtual operator, and infrastructure investments may be used to deter entry. If the cost of investing is sufficiently convex, then the equilibrium level of investment turns out to be low and the investing firm will accommodate entry by the virtual operator.

We only consider the case where the facility-based firms choose accommodation of entry by the virtual operator. In order to ensure this we assume that the costs are sufficiently convex (*i.e.* $\varphi > 2$). With accommodation of entry the results obtained in Proposition 3 hold. For lower values of φ, it may be optimal for the facility-based firms to invest to foreclose the virtual operator from the market. In this case, the higher the

β^v compared to β^f, the more the facility-based firms have to invest in order to deter entry.[22]

When roaming is mandatory, a welfare-maximizing regulator decides the appropriate levels. The regulator will choose a roaming policy that corresponds to the voluntary roaming case. This seems to correspond well to intuition. By keeping β^v low and β^f high, firms have better incentives to invest in infrastructure. A higher level of investment again leads to more output being produced in the final product market, which is a direct benefit to consumers. Consequently, there is little scope for regulatory intervention in this case. The main result in this section suggests that the provision of incentives to invest be of greater importance than allowing the virtual operator entry at equal terms to the incumbents. Allowing entry at equal terms will increase competition and output, but this effect is dominated by the effect through the investment incentives.

2.2. Incumbents set the investments cooperatively

When the facility-based firms set investments cooperatively they maximize the joint profit, and by inserting for optimal stage 3 quantity we obtain the following equilibrium investment level:[23]

$$x = \frac{(a-c)(1+\beta^f-\beta^v)}{4\varphi - 2(1+\beta^f-\beta^v)^2} \tag{13}$$

Hence, we have the following results:

Proposition 4: When facility-based firms invest in infrastructure cooperatively, the investment level is unambiguously increasing in the degree of roaming quality between the cooperating firms, $\partial x/\partial \beta^f > 0$, and decreasing in the roaming quality between facility-based firm and the virtual operator, $\partial x/\partial \beta^v < 0$.

The proof is straightforward and hence omitted.

Similar to our findings when investment is undertaken non-cooperatively, we observe that the equilibrium investment level is reduced whenever the roaming quality between the investing firms and the virtual operator increases. The intuition is the same as in the non-cooperative case. One important reason for investing in infrastructure is to differentiate its product from that of the competitor. If the roaming quality between the investing firms and the virtual operator is sufficiently high, the importance of the differentiation effect diminishes. The roaming quality may be interpreted either as the quality of the access provided to the virtual operator or as the virtual operator's capability

22 In a similar context, the case where a facility-based firm over-invests to deter entry is analyzed by Foros (2000).

23 Note that for the equilibrium investment level to be positive, the investment cost needs to be sufficiently convex; this requires $\varphi > 2$.

of transforming the product innovation resulting from the investment made by the facility-based firms into high quality final products. Consequently, our findings suggest that any increase in the quality of input or in the virtual operator's capabilities reduces the facility-based firms' incentives to invest in infrastructure.

Our results suggest that there is little scope for regulation when investments are undertaken cooperatively both with voluntary and mandatory roaming. This is also the case when investments are undertaken non-cooperatively. Both the investing firms' interests and the interests of a welfare-maximizing regulator coincide. When the facility-based firms cooperate at the investment stage, they maximize joint profit with respect to (β^f, β^v), whereas a welfare-maximizing regulator chooses (β^f, β^v) to maximize the sum of profits and consumers' surplus. The solution in both cases yields maximum roaming quality between the facility-based firms and minimal roaming quality between facility-based firms and the virtual operator. In order to ensure non-negative profits for the virtual operator when $\beta^v = 0$ we must however make a stronger restriction on the convexity of the investment cost function. A sufficient condition for non-negativity is: $\varphi > 4$.

For a given investment level, an increase in the degree of roaming quality to the virtual operator is a social benefit and adds to consumers' surplus. However, increasing the roaming quality to the virtual operator adversely affects the facility-based firms' incentives to invest in infrastructure, which is detrimental to consumers' surplus. When a virtual operator is allowed to enter and investments are undertaken collusively the trade-off is the same as when investments are undertaken non-cooperatively, and the investment incentives dominate.

2.3. *Some remarks on entry of a virtual operator*

One remaining question is whether competition authorities (or regulators) should allow a virtual operator to enter or not. It is reasonable to assume that such entry should be encouraged if the entry implies that welfare is higher than in the absence of the virtual operator. We assume that the decisions to allow entry by a virtual operator and whether to allow collusion are taken prior to commencement of the three-stage game analyzed above. Without going into details, it can be shown that welfare is higher if collusion is allowed at the investment stage (as is also the case without a virtual operator), and the reason is that the positive externality can be internalized more easily under collusion.

Since we know that welfare under collusion is higher than in the non-cooperative case, we need to show that welfare under collusive investments with a virtual operator is higher than without the virtual operator. It can be shown that welfare is indeed higher when entry of a virtual operator is allowed and there is collusion at the investment stage. This implies that the subgame-perfect equilibrium policy for the government is to allow entry by a virtual operator and allow facility-based firms to cooperate at the investment stage.[24] We have seen that the optimal roaming policy implies that the virtual operator is only

24 The proof of this result involves messy, but straightforward algebraic manipulations.

given minimal roaming quality, which means that the entry for the virtual operator is not on particularly generous terms.

A final point to be made is that the introduction of a virtual operator can eliminate the commitment problem for firms. Note that firms choose the roaming quality prior to undertaking investments, and in the absence of a virtual operator firms may have incentives to change the quality of roaming after investments have been sunk. When a virtual operator is allowed to enter, this commitment problem is eliminated, and firms have no incentives to change the quality of roaming after investments are sunk.

3. Concluding remarks

We have discussed roaming policy (both voluntary and mandatory), and we have also briefly discussed some competition policy aspects related to sharing infrastructure in the mobile communications market. In particular, we have focused on the interaction between roaming policy and investment incentives in the third generation mobile networks (*e.g.* UMTS). We have shown that all involved are better off under collusion provided that roaming quality is set sufficiently high. Furthermore, in our model, the chosen level of roaming quality is indeed sufficiently high in all cases. This implies that when a regulator or competition authority chooses whether collusion at the investment stage should be allowed, they know that in whatever policy they choose with respect to roaming they can never do worse than allowing collusion. In some of the Nordic countries, major players in the mobile communications market have decided to cooperate in the process of setting up the next generation mobile networks, which could be interpreted as the collusion case in our model. In danger of stretching our model too far, we have shown that such collusion is actually beneficial in terms of welfare. This implies that competition authorities should not interfere with such cooperation.

When we introduce a virtual operator into the game that relies on the facility-based firms for infrastructure access, we find that the relationship between roaming quality and investment incentives is qualitatively different. Our findings also suggest that there is little scope for regulation of roaming quality when there is a virtual operator present, both under cooperative and non-cooperative investments. This is also different from the case without the virtual operator present, where the social optimum does not correspond to the unregulated outcome if investments are undertaken non-cooperatively.

References

Brod, A. and Shivakumar, R., ''Advantages of semi-collusion,'' The Journal of Industrial Economics, vol. 47, pp. 221–230, 1999.

Busse, M.R., ''Multimarket contact and price coordination in the cellular telephony industry,'' Journal of Economics & Management Strategy, vol. 9, pp. 287–320, 2000.

Cave, M. and Valletti, T.M., ''Competition in the UK mobile communications,'' Telecommunication Policy, vol. 22, pp. 109–131, 1998.

Ceccagnoli, M., ''The effects of spillovers and rivalry among firms with heterogeneous technological capabilities on R&D efforts,'' Mimeo, 1999.

Crémer, J., Rey, P., and Tirole, J., ''Connectivity in the commercial internet,'' The Journal of Industrial Economics, vol. 48, pp. 433–472, 2000.

d'Aspremont, C. and Jacquemin, A., ''Cooperative and noncooperative R&D in duopoly with spillovers,'' American Economic Review, vol. 78, pp. 1133–1137, 1988.

De Bondt, R., ''Spillovers and innovative activities,'' International Journal of Industrial Organization, vol. 15, pp. 1–28, 1997.

Fershtman, C. and Gandal, N., ''Disadvantageous semicollusion,'' International Journal of Industrial Organization, vol. 12 no. 2, pp. 141–154, 1994.

Foros, Ø., ''Strategic investments with spillovers, vertical integration and foreclosure in the broadband access market,'' Manuscript, 2000.

Kamien, M.I., Muller, E., and Zang, I., ''Research joint ventures and R&D cartels,'' American Economic Review, vol. 82, pp. 1293–1306, 1992.

Katsoulacos, Y. and Ulph, D., ''Endogenous spillovers and the performance of research joint ventures,'' The Journal of Industrial Economics, vol. 46, pp. 333–357, 1998a.

Katsoulacos, Y. and Ulph, D., ''Innovation spillovers and technology policy,'' Annales d'Economie et de Statistique, no. 49–50, pp. 589–607, 1998b.

Katz, M., ''An analysis of cooperative research and development,'' Rand Journal of Economics, vol. 17, pp. 527–543, 1986.

Kreps, D.M. and Scheinkman, J.A., ''Quantity precommitment and bertrand yield cournot outcomes,'' Bell Journal of Economics, vol. 14 no. 2, pp. 326–337, 1983.

Leahy, D. and Neary, J.P., ''Public policy towards R&D in oligopolistic industries,'' American Economic Review, vol. 87, pp. 642–662, 1997.

Matthews, J., ''Virtual mobile services: Strategies for fixed and mobile operators,'' London, Ovum (Ovum reports), 2000.

Parker, P.M. and Röller, L.-H., ''Collusive conduct in duopolies: multimarket contact and cross-ownership in the mobile telephone industry,'' Rand Journal of Economics, vol. 28, pp. 304–322, 1997.

Poyago-Theotoky, J., ''A note on endogenous spillovers in a non-tournament R&D duopoly,'' Review of Industrial Organization, vol. 15, pp. 253–262, 1999.

Salant, S.W. and Schaffer, G., ''Optimal asymmetric strategies in research joint-ventures,'' International Journal of Industrial Organization, vol. 16, pp. 195–208, 1998.

Steen, F. and Sørgard, L., ''Semicollusion in the Norwegian cement market,'' European Economic Review, vol. 43 no. 9, pp. 1775–1796, 1999.

Biography

Øystein Foros is a Ph.D. student at the Norwegian School of Economics and Business Administration (NHH) and a Research Scientist at Telenor R&D. His main research areas include industrial organization, regulation, and the economics of information and communication technology (ICT).

Bjørn Hansen is a Research Manager at Telenor R&D. His main research areas include management and business strategy in telecommunication markets.

Jan Yngve Sand is an Associate Professor in Economics at the University of Tromso, Norway, and a Research Scientist at Telenor R&D. His main research interests include regulation, industrial organization, and public economics.