

MPRA

Munich Personal RePEc Archive

Mobile Call Termination: a Tale of Two-Sided Markets

Tommaso Valletti

Communications & Strategies

March 2006

Online at <http://mpa.ub.uni-muenchen.de/2605/>
MPRA Paper No. 2605, posted 6. April 2007

Mobile Call Termination: a Tale of Two-Sided Markets (*)

Tommaso VALLETTI

Imperial College London, University of Rome "Tor Vergata" and CEPR

Abstract: Mobile telephony is described as a "two-sided" market where customers are seen as senders and receivers of communications that are mutually beneficial both to callers and receivers. This has implications in terms of market definition and market power. The economics of mobile call termination is discussed in this context.

Key words: mobile telephony, market definition and call termination

■ Market definition in mobile telephony

The standard test adopted by most anti-trust and regulatory authorities to identify markets is the so-called SSNIP test (sometimes also referred to as the "hypothetical monopolist test"). This is designed to explore the consequences of a (hypothetical) Small but Significant and Non-transitory Increase in Price on the profitability of the (hypothetical) firm that initiates it. At the heart of this test lies the question of what might make such a price rise unsustainable. Some consumers may switch to substitute products ("demand-side substitutability") and some firms operating "near" to the (narrowly defined) candidate market may alter their plans and supply similar products ("supply-side substitutability"). If there are close demand- or supply-side substitutes, then the price increase initiated by the hypothetical monopolist will lead to a large reduction in its sales, and its profits will, as a result, fall.

(*) Parts of this paper are based on work carried out by the author for the European Commission. The comments from an anonymous referee are gratefully acknowledged. The opinions expressed in this paper are the sole responsibility of the author.

A number of difficulties arise in identifying market boundaries including deciding how to treat firms that operate in many related markets, dealing with intermediate goods markets, applying the test to markets that are already monopolised (known as the "cellophane fallacy"), and determining what is "small but significant".

All of these difficulties occur when applying these general principles to mobile telephony markets¹. Customers buy mobile telephones for many reasons. Customer profiles are extremely heterogeneous in terms of calling patterns, needs, mobility, etc., which is partly reflected in the vast number of tariffs on offer in these markets. The needs of a certain customer are themselves not immutable, and will depend on factors such as circumstances and locations. In principle, therefore, if one defined an antitrust market in a very narrow way and purely on the basis of substitutability at a given point in time, this exercise would result in a proliferation of very narrowly defined markets. At the same time, however, a mobile operator is a provider of different products and services that satisfy these various needs. In other words, a mobile operator can be seen as a multi-product firm. The fact that a firm manufactures or sells more than one product may suggest, but by no means implies, that there should be a much bigger market for that firm's total output. According to this view, the relevant market should include a "cluster" of products, where non-substitutes should be included in the same market.

The concept of cluster markets clearly applies to most services in mobile telephony. Customers typically want one handset and one SIM card to handle almost all their calls, SMS, etc. Even if one accepts the broader concept of a cluster market, an extra layer of complications arises in the context of mobile telephony because benefits and costs associated with calls generally do not accrue to the same party. When a conversation happens, there must be both "senders" and "receivers" involved, which are, by definition, different individuals. Clearly, no one would ever want to place a call if that call is known not to be received or ever retrieved. Even more obviously, one cannot receive a call if this call has not been made! As obvious as this may sound, it is a healthy reminder of the type of economic considerations that must be taken into account when defining markets, without risking the derivation of fictional market definitions.

¹ This paper deals with mobile telephony only, although most of its arguments are also valid more generally, including for a deregulated fixed telephony sector. I prefer to stick to the mobile case to avoid crucial factors specific to fixed telephony, such as extremely large incumbency advantages, public ownership or universal service obligations.

As an example, it is a common and useful practice to think of a retail market for call origination, although it is clear that this market cannot exist in isolation without termination. When the SSNIP test is applied to the market for call origination, the analysis should try to assess how the call originator would respond to an increase in price, looking for possible substitute services etc. This analysis presupposes that the same change in demand for calls originated by the sender will also occur on the receiving side, i.e., every call is accepted by the receiver. This is, indeed, a very likely situation since receivers will not pay for the call in most cases. According to this line of analysis, the retail market for call origination is de facto extended to include termination as a necessary input for an originated call to be completed. Termination is an input that is not directly bought by the call originator, but is needed to satisfy the call originator's needs.

According to this view, there is a retail market for call origination, but not a retail market for termination, which is a derived demand instead. Call origination and call termination are in a vertical relationship where the provider of call origination takes as given the input price for termination, and then charges a mark up depending on the price elasticity of outgoing calls. A market analysis could therefore find that the retail market for call origination is competitive, but the input market for termination is monopolised (and vice-versa). The distinction between call origination at the retail level and call termination at the wholesale level is, to a large extent, fictitious and merely reflects common billing practices, rather than the underlying economic vertical relationship in the production of a (completed) telephone call (see box 1 below).

Box 1 – Termination: retail or wholesale market?

Imagine customer A calls customer B and pays p_{AB} to A's provider. A's provider then pays a termination charge t_B to B's provider. The competitive environment that leads to the setting of p_{AB} (at the retail level) may have nothing to do with the competitive environment that leads to the setting of t_B (at the wholesale level). Alternatively, imagine a situation where there is no inter-carrier compensation, and customer A pays directly p_A to provider A for call origination and p_B to provider B for call termination. In the eyes of customer A, the two situations are formally equivalent if, for instance, $p_A = p_{AB} - t_B$ and $p_B = t_B$. Once again, the competitive conditions that lead to the setting of p_A and p_B (both at the retail level under this alternative pricing arrangement) could be very different.

The previous example is coherent, but incomplete. In fact, we argued that a market for call origination can only exist if there is also a market for termination. Implicitly in the previous lines of argument, we assumed that termination was needed only by the sender. However, if a call is accepted by a receiver, then this implies that there is also a demand for termination of

calls on the side of the receiver! If one then applies the SSNIP test to this market, the exercise looks less straightforward. Which price should one increase? And who pays for it? The response of a customer to an increase in the price of termination, and therefore the profitability of the (hypothetical) firm that initiates it, will differ if the party that bears its cost is the receiver or the sender.

A less formal market definition would at this stage consider the whole economic environment, starting from the fact that customers do not demand calls per se, rather they want to communicate, for example, exchange information. Calls sent and received are just inputs in this exchange of information. According to this view, a mobile operator is a provider of a "platform" that allows the exchange of communications between these two different sides, the senders and the receivers. In this sense, a mobile firm should be analysed in the context of the "two-sided markets" framework, which has recently received much attention both in academic literature and in court cases.

■ Two-sided platforms

The term "two-sided platforms" (2SPs) refers to products and services that must be used by two (or more) groups of customers to be of value to them. The "platform" enables interactions between the different "sides", trying to get the two sides "on board", and charging each side.

2SPs are the subject of a recent body of academic literature in economics that usually refers to them as "two-sided markets"². Since the term "market" is used in a different way for the purposes of antitrust policy, this paper adopts the more neutral 2SP terminology³. There is no unequivocal definition of 2SPs in existing literature. Rochet and Tirole (2003) proposed the following definition: "A market is two-sided if the platform can affect the volume of transactions by charging one side of the market more and reducing the price on the other side by an equal amount; in other words, the price structure matters".

² See ROCHET & TIROLE (2003), EVANS (2003), WRIGHT (2004), ARMSTRONG (2006).

³ See EVANS & NOEL (2005).

The previous definition draws an important distinction between price structure and price level. This makes 2SPs different from markets encountered in textbook economics, where the price structure is typically neutral. For instance, in competitive markets it is irrelevant who is charged VAT, whether this be the producer or the consumer, since only the price level matters for the level of transactions between the two sides (buyers and sellers). In 2SPs, on the other hand, the price structure that the two sides are charged has an impact on allocation. If the two sides cannot internalise externalities between them, then the Coase theorem does not apply and market failures can arise. The role of the platform can therefore be that of an intermediary, finding the right pricing structure between the two sides and allowing trade to take place.

An alternative definition immediately follows from the previous discussion. A 2SP arises in a situation where: (a) there are two (or more) sides, with (uninternalised) inter-group network externalities, and (b) platforms have the ability to price discriminate between the two sides.

Definitions aside, it is helpful to give a few examples of 2SPs. EVANS (2003) introduces a useful taxonomy of 2SPs:

- *Exchanges* such as security exchanges, auction houses, brokers, and various matchmaking activities (for example, employment agencies and real estate agents). Exchanges help buyers and sellers search for feasible contracts. The externality here arises from the fact that having large number of participants on both sides increases the probability that participants will find a match.

- *Advertising-supported media* such as newspapers, directories, television, and web portals. Media provide contents that attract audiences. Audiences, in turn, are used to attract advertisers. There are two kinds of externalities between the two sides. Audiences exert a positive externality on advertisers, as advertisers value platforms that have more viewers. On the contrary, advertisers exert a negative effect on viewers, at least to the extent that commercials interrupt a programme, or make it more difficult to consume content.

- *Transaction systems* such as credit cards. These are similar to exchanges in some respects, as cardholders and merchants are more likely to adopt a particular credit card the greater the number of adopters of the same card on the other side. They also have some peculiar features, namely card associations are cooperative 2SPs: for a transaction to be completed there must be an agreement as to the division of profits and the allocation of

various risks between the entity that services the cardholder and the entity that services the merchant.

- *Software platforms* such as PCs, video games and music players. The two sides here are represented by users who want to run software applications and developers who write applications and sell them to users.

Are 2SPs relevant for telephony? Clearly, any network operator is a multi-product firm. However, the mere fact that multiple product or "cluster" markets are involved does not imply that a 2SP is implicated. If the various products are bought and consumed by the same customer, there is no 2SP involved since there are no inter-group network externalities. Therefore, services such as access and call origination can be analysed, to a large degree, with standard antitrust tools that do not need to be extended to the analysis of 2SPs.

There are situations where 2SPs can be applied to telephony too. An important case in point is call termination. A network operator, in this case, falls in the category of "exchanges" introduced above, as it allows "senders" and "receivers" to complete their match, i.e., communicate. There is an externality involved as senders can communicate more the higher the number of receivers they can contact, and receivers are likely to benefit from receiving many calls the larger the number of senders there are⁴. More generally, termination revenues form an integral part of the way an operator sets prices for both termination and outgoing services. These can be distinct services, but have close inter-relationships since the demand and price for one service affects the other.

Although we will analyse call termination markets only in a later section, we anticipate here that the exercise of market power when setting termination rates is likely to differ when calls are sent and received "on-net" (i.e., senders and receivers both subscribe to the same network operator) and when they are "off-net" (i.e., senders and receivers belong to different network).

In the former case, the "platform" is likely to internalise externalities between the two sides, and the presence of competition limits the ability of

⁴ It could be argued that mobile users belong to the same group. One should therefore speak of "intra-group externalities", rather than "inter-group" externalities typical of 2SPs. However, please note that my description of the problem relies on having "senders" and "receivers", which represent the two groups that need a platform to conduct an exchange of communications. In this sense, I would argue that the definition of a 2SP applies to mobile telephony literally.

the network operator to raise termination prices. In the latter case, the network operator will not internalise the effects on senders when setting the termination rate and market failure is likely to arise. A specific example of such market failure is the case of fixed-to-mobile (F2M) calls ⁵.

Two-sided platforms: market definition and market power

When applying market definitions to 2SPs one has to be particularly careful to avoid mechanical applications of commonly used concepts due to the possibly intricate relationship between the various sides. When dealing with a 2SP, it is essential to evaluate if network effects (i.e., links between the two sides) are: (a) present, and (b) limit the extent to which a price increase on either side is profitable. This exercise is tricky as it mixes several factors: which price should be increased? Who pays for this increase? What is the starting level for the price increase? Should a firm re-adjust its entire structure of prices when only one price changes?

Take, as an example, the case of F2M calls and mobile access. Are they complements or substitutes? The answer to this type of question is of some use in "normal" markets, as substitute goods are typically presumed to belong to the same relevant market. Imagine first an increase in the price of mobile access. Demand for mobile access would go down as a direct consequence of the price increase. As there would then be fewer mobile customers to call, demand for F2M calls would also fall. As seen from this perspective, F2M calls and mobile access seem to be complements. Now imagine increasing the price of mobile termination, starting from the termination cost. Demand for F2M calls would decrease because fixed users would have to pay more to call mobile phones. However, the increase in the price of mobile termination has also introduced some termination revenues that did not exist when termination was set at its cost. If there is some competition for mobile users, these termination profits should, at least to some extent, be passed on to mobile users. A likely scenarios would be for the mobile network operator to push down the price of mobile access. This, in turn, should boost demand for mobile access. From this point of view, an increase in the price of F2M termination increases demand for mobile access, while F2M calls and mobile access now seem to be substitutes!

⁵ The theory of two-sided markets received some prominence in the recent case on mobile termination rates in New Zealand; see NZ Commerce Commission (2005).

It is beyond the scope of this paper to conduct a full analysis of the termination problem here⁶. Our main point is that questions, such as whether F2M calls and mobile access are complements or substitutes, do not make much sense when they mechanically apply standard notions of substitutability and complementarity to highly specific market realities, such as 2SPs. As we have already seen, a mobile operator is a "platform" that provides access among other things (and the corresponding price is paid by mobile consumers), but also enables the termination of calls initiated by fixed users. The price for termination is indirectly paid by fixed users and, typically, not by mobile users. These are the main features that have to be taken into account when conducting an economic analysis of the termination problem.

Another important caveat, when defining markets in the presence of 2SPs, applies to the use of the SSNIP test. Firstly, when a price is increased, the corresponding demand will decrease, as in standard markets, but there may also be additional effects arising from the other side that may, or may not decrease the profitability of the price increase, according to the type of inter-group network externalities involved. For instance, in an exchange such as a matchmaker, where one side benefits from the presence of high numbers from the other side, imagine the platform increases the price it charges to one particular side. This will reduce the number of buyers from this side, making it less appealing for the other side to join the platform, further reducing demand from the original side. In this case there is a "multiplier" effect, as a price increase reduces demand more than in standard one-sided markets. In the case of advertising-supported media, on the other hand, imagine the platform increases the price it charges one side (advertisers). This should decrease the number of commercials bought by advertisers, making it more appealing for the other side (viewers) to join the platform⁷. Secondly, it is not clear where the hypothetical price increase should originate from. The cost of a product is typically not an efficient benchmark in the presence of 2SPs. Perhaps more disappointingly, even the price level set in a "competitive market" is not efficient. This should not come as a surprise since it is well known in economics that competitive markets "work", i.e., they are efficient and any intervention could just make things worse, only without externalities. This

⁶ See WRIGHT (2002), and VALLETTI & HOUPIS (2005).

⁷ I assume that other variables such as programme quality or content are not affected. The important point here is that it is easy to construct situations where the "multiplier" effect can go either way.

fundamental result can be rephrased by saying that, in the presence of externalities, even competitive markets do not work and some appropriate intervention can increase the welfare of society.

It therefore seems that trying to define sharp boundaries can be a risky exercise with 2SPs. Since from a legal standpoint, in practice, market definition requires that a product is found to be either in the market or outside it, a possible reasonable compromise would be to look at standard (possibly narrow) market definition to start with. Then, the impact on competition in "affected" markets (therefore, extending the analysis beyond the original market definition) could be considered at a later stage when conducting a full economic analysis, eventually leading to the imposition of appropriate remedies. Alternatively, one could start with the whole products under consideration, avoiding the exercise of market definition and directly delving into the economic problem at stake. For an economist, this second approach is bound to give the same answer (and therefore the same set of possible remedies) as the first approach. However, it is not clear if, from a legal standpoint, these two approaches are also identical. For instance, SMP may be found over the narrowly defined market, which would imply the introduction of some remedies, "adjusted" for the two-sidedness feature of the market investigated. However, SMP may not be found if one started from the whole set of interlinked products, where SMP is linked to the presence of some extra rent that the firm can sustain overall. Therefore an investigation may not start although "welfare enhancing" regulation would also be available in this case.

While there is not much disagreement on economic analysis, there may be some divergence between the legal and economic approach to the main questions addressed. This is a fundamental and controversial point that harks back to the meaning of SMP and the ultimate objective of regulatory and antitrust intervention. Competition law can maintain competition, but typically cannot create it or cure defects or market failures. It also cannot impose very precise obligations. On the contrary, regulation usually has aims that are wider than those of competition law, and has methods that go beyond those of competition law, because regulators can impose additional or new duties necessary to promote the objectives specified. In the specific context of 2SPs, it therefore follows that competition law should not be able to deal with inefficient pricing structures arising from competition in two-sided markets. This is because competition law assumes that firms can unilaterally desist from the conduct that is undesirable. Fines and other anti-trust sanctions rely on firms being able to take unilateral action to comply and to act competitively. However, in 2SPs a firm cannot unilaterally lower a

particular price that is deemed to be "wrong" (for example, too high) if the other competitors do not - that would result in losses relative to the rivals. The threat of fines thus does not work in this context, because no individual firm can comply. The consequence of this reasoning is that any intervention has to ensure collective compliance - either by all firms having the same unilateral incentives at the same time (for example, by setting up a position in which the authority effectively requires them to "collude"!) or by their conduct being subject to some exogenous constraint (which is another word for regulation).

Conclusions on 2SPs

2SPs involve inter-group network externalities and are relevant in many industries, including telecommunications. As a result of these externalities, socially-optimal prices in 2SPs typically depend in some intricate way on price elasticities of demand, inter-group network effects and costs. This is a complex exercise that can be conducted by taking into account market realities and avoiding mechanical applications of standard definitions and tools.

Another result of externalities is that socially-optimal prices in 2SPs, generally, are not purely cost-based. By understanding the nature of the problem, it is therefore easy to avoid possible fallacies. For instance, incremental cost pricing is typically not efficient with 2SPs. High individual mark-ups may also not indicate standard market power. A more balanced pricing structure (interpreted as prices being more in line with costs) is not necessarily produced by fiercer competition. Moreover, the removal of alleged cross-subsidies, such as decreasing one price (A) and increasing another price (B), does not necessarily benefit the side (A) that pays a price above cost. This is because, by increasing the other price (B), some B users may drop off, thus making the product less valuable to A users as well.

Firms with the features of a 2SP are correct to stress the fact that these are special markets, which policy-makers consequently need to be very careful with. We agree with this point and always advocate a full and appropriate economic analysis of these markets. However, we conclude by recalling that, even if a two-sided market is assumed to be perfectly competitive, then the market does not work. This is in stark contrast with standard one-sided markets: when these markets are competitive, they are also efficient and no regulator should interfere with their working. In two-sided markets, on the

other hand, privately chosen prices, even when ideally set by competing firms, will differ from socially-optimal prices. An appropriate intervention can increase consumer and social welfare. 2SPs should therefore be subject to more, rather than less regulatory oversight.

■ Incoming and outgoing calls

Let me now return to market definition in mobile telephony. People buy mobile phones to have access, that is, what they buy is the ability to make and receive different kind of calls while travelling in different places. Access typically involves the purchase of a handset and a SIM card. After having secured access, customers then use their phones, that is, they do make and receive different kinds of calls while travelling in different places. Access, outgoing calls, and incoming calls are the three general groups of services that represent the starting point of the analysis of market definition in mobile telephony. One consequently needs to understand how a customer would react when a hypothetical monopolist increases the price of one of these three services.

This apparently simple exercise has to be done while taking into account relevant features of the economic environment under consideration. A crucial aspect in the mobile telephony industry is that, in the absence of any intervention, the party making and paying for the call is typically the sender and not the receiver of the call. This arrangement, known as CPP ("Calling Party Pays") is adopted in all countries in the EU. Under CPP, the service is initiated by, and paid for by, the caller to the mobile phone, not the mobile phone owner. A SSNIP test conducted on the price of access or outgoing calls is therefore a very different exercise to a hypothetical increase in the price of incoming calls, since, under the current pricing arrangements, in the former case it is the telephone owner that pays directly for the price increase, while in the latter there is no direct payment involved, although the receiver may indirectly suffer from receiving less calls. As a result of this fundamental difference, the analysis of access and outgoing calls has to be kept separate from the analysis of incoming calls. This paper does not consider access and outgoing calls here, as the analysis can be conducted, to a large extent, with standard tools (such as "cluster" markets), but focuses instead only on the market for "incoming" calls.

Incoming calls

Mobile customers want to receive calls. Under the CPP system, these calls are initiated and paid by other customers. Given this peculiar feature, the exercise of market definition should be conducted looking at the behaviour of both the sender and the receiver. Let us start with the sender first. The sender has a demand for calls to a particular person owning a mobile phone. Calls to mobile phones do not have strong demand substitutes, as senders typically are willing to pay a premium if they need to contact a person without knowing her exact location. If the price of a call to a mobile network goes up, a caller would probably reduce the number and/or length of calls, according to her demand elasticity, but it is very unlikely that the caller can find good alternative substitutes. A call is typically placed to a mobile user when the caller wants to be sure to contact and interact in real time with the called party, for which there is no effective substitute. The sender therefore has very limited ability to find substitutes if the price of calls to mobile goes up because of a price increase initiated by the mobile operator that terminates the call ⁸.

The behaviour of senders therefore does not impose any limit on the ability of the mobile firm to increase the price of incoming calls. However, this analysis is incomplete since constraints on increases in the price of incoming calls can also arise if receivers themselves react to an increase in the price of a call to a mobile. For instance, if the receiver cares about the satisfaction of the sender, then the price of calls to mobile telephones will be internalised. The latter case is sometimes referred to as "closed user groups" and can correspond to families that behave under a single budget constraint, or some business users who provide different sorts of telephony services to their employees. These can constitute a large part of the customer base of a mobile operator. Mobile operators, however, have the ability to price discriminate among different groups, for instance by offering discounts to large business users, hence their presence does not seem to restrict overall price levels for other customers.

⁸ Continuing with the example presented in box 1, where customer A is the caller and customer B is the receiver, this price increase could be paid directly by the sender if the price p_B for termination is paid directly by customer A to B's provider at the retail level. If, instead, A's provider bills customer A and then pays a termination charge to B's provider, the price increase would be initiated at the wholesale level (t_B) and have repercussions at the retail level (p_{AB}). In this latter case (the most common situation in practice), the demand for B's provider is a derived (input) demand to be analyzed at the wholesale level. In both cases, however, customer A has a limited ability to find a substitute means of contacting customer B.

The receiver may still limit the provider's ability to charge others high prices. In fact, if the price of incoming calls increases, the number of calls received will decrease, which has a negative effect on the satisfaction of the receiver, since receiving calls is clearly one of the incentives of subscribing to a mobile telephone in the first instance. However, this is not necessarily a disadvantage for consumers that receivers can easily see or react to. It is documented by several NRAs (for example, Ofcom) that receivers' awareness of the price of calls to mobile telephones is low and that the price of incoming calls is not considered by subscribers to be an important factor in their choice of mobile operator and other factors are more influential. The mobile owner cares most about the prices s/he has to pay to subscribe to and place calls with a mobile operator, but in most cases will not take into account the prices paid by other callers to contact him/her. In fact, mobile telephone owners may enjoy a higher level of overall satisfaction if an increase in the price of incoming calls, despite reducing the number of incoming calls, induces the mobile operator to decrease other prices directly paid by subscribers.

When assessing what type of dominant behaviour might arise in the market for incoming calls, it is useful to distinguish between the following three types of mobile incoming calls:

- calls to mobile (on-net),
- calls to mobile (off-net),
- calls to mobile (from other non-mobile networks, mostly F2M calls in practice).

In principle, given that a mobile firm is by definition the only firm that can terminate calls for its own customers, SMP in the form of single dominance should arise, no matter what type of call is under consideration. However, as mentioned repeatedly above, in this market both a sender and a receiver are involved and their identity cannot be neglected.

In the case of on-net calls to mobile, if the mobile firm tried to increase the price of the termination end of the call, the sender that would suffer the price increase would be one of its own customers. An increase in termination price would make the overall package offered by the firm to its subscribers less appealing, and the firm would lose customers as a result. Competitive forces do act as a constraint on the firm's behaviour, hence there is not likely to be any market power abused in this case. In terms of the analogy with two-sided markets, in this case the mobile firm is a platform that perfectly "internalises" transactions that only affect its customers.

Contrary to on-net calls, single dominance is likely to exist for the other two kinds of incoming calls, mobile off-net calls and calls to mobile from other networks (F2M calls). In these two instances, the sending party that pays the call is not one of the firm's customers and the firm's receiving customers would not react to a price increase, which gives the mobile firm the ability to set the price at monopoly levels. From the point of view of single dominance, these two types of calls are therefore quite similar.

There is nonetheless one possible important difference between these two types of incoming calls to mobile from other customers. The difference lies in the strategic environment. Off-net calls are charged to customers belonging to a rival mobile network, while there no strategic interaction between a mobile firm and a fixed firm, as these are to a large extent separate markets.

As customers buy mobile phones with the purpose of receiving calls from other customers, a firm might be tempted to increase its off-net termination price in order to distort competition in the market. This incentive exists, on top of the termination monopolisation effect, only for mobile off-net calls. For instance, a mobile firm could set a high off-net termination charge, so that the overall off-net price paid by rival customers is high. Customers would be willing to join a bigger network: on-net calls, to the extent that they are cheaper than off-net calls, imply that customers would be receiving relatively more incoming calls.

What we have described so far can be said about the price incoming calls in general, without distinguishing whether these calls are set at the "wholesale" level as termination charges or at the "retail" level charging senders directly (see box 1 again for this analogy). There is, however, a possible main difference with the "retail" market analysis of incoming calls. If the sending party was billed directly by the receiving operator, it seems natural that the termination price is set directly by the receiving network, thus the sending customer has no bargaining power. Instead, at the wholesale level, the termination price is more likely to be negotiated between the sending and the receiving network. Countervailing buyer power (i.e., bargaining, negotiations) should therefore be taken into account when analysing the wholesale market for incoming calls in order to determine the presence of SMP.

In particular, a bargaining model seems quite appropriate to an analysis of the market for "off-net" M2M calls, as this is a bilateral problem of "two-way" interconnection, where two wholesale prices have to be negotiated,

one in each direction. One network, when negotiating the wholesale price for sending calls to the rival network, can always use its own wholesale price for receiving calls from the rival as an effective threat in the bargaining game. In this context, there are different sets of results from the literature ⁹:

- Bilateral wholesale negotiations can get rid of inefficiencies, given the reciprocal nature of bargaining. This is true particularly for negotiations among symmetrically-placed networks.

- Bilateral negotiations may be used to affect the intensity of competition at the retail level. The nature of collusion may be different:

- Collusion may happen in a "static" framework by setting high termination rates because of a "raise-each-other's-cost" effect ¹⁰. This result holds true only under particular circumstances, namely retail prices should be linear (which may be applicable to pre-paid cards), while it does not apply under more sophisticated retail pricing structures (two-part tariffs, for example, monthly rental plus price per minute of usage).
- Collusion may also happen in a more standard "dynamic" framework, where networks repeatedly interact with each other. The role of wholesale termination charges may be one of giving a "focal" reference point to set collusive retail prices. Please note that, in this case, joint dominance should be established at the retail level, while the wholesale level may facilitate reaching the collusive agreement.

The applicability of a bargaining model to the determination of the wholesale price for termination of F2M calls is more controversial ¹¹. In a bargaining model, two parties have to find a way to divide the surplus created by finding a deal. This division is influenced by the outside options that the parties have, i.e., what they could get if they threaten not to strike a

⁹ See ARMSTRONG (2002), LAFFONT & TIROLE (2000), VOGELSANG (2003), CAMBINI & VALLETTI (2005).

¹⁰ To see this, imagine what happens when operators charge customers collusive (monopoly) retail prices. If mobile customers call each other with the same probability, the traffic is balanced and an operator pays the rival the same amount for termination services that it receives from the rival for similar services, independently of the value taken by the termination charge. This can be an equilibrium only if no one has a unilateral incentive to deviate. If one firm deviates from the monopoly retail charges by undercutting the rival, it induces its subscribers to call more. Since part of the calls made is destined for the rival's network, the effect of a price cut is to send out more calls than it receives on-net from the rival. The resulting net outflow of calls has an associated deficit that is particularly burdensome if the unit termination charge is high. This will discourage under-pricing in the first place. Some conditions are necessary to produce this outcome, for instance products should not be too homogeneous, otherwise the incentive to undercut would have the additional benefit of increasing market share.

¹¹ See BINMORE & HARBORD (2005), UK Competition Appeal Tribunal (2005).

deal. The "threat" points are not as natural as in the bilateral negotiation of termination of M2M calls. In the case of F2M calls, the negotiated price is only "one way", as the other way (M2F) is typically regulated. This asymmetric treatment of M2F and F2M calls is a possible source of distortion that must be noted.

This problem of "bargaining in the shadow of regulation" still has to be clarified in full. However, some related aspects have received partial answers. For instance, an argument put forward has been that, to have a viable business, a small MNO must have an interconnection agreement with the incumbent fixed-network operator. This argument mixes up incoming calls and all other services. In fact, as a first cut, the bargaining problem does not seem to be affected by the size of a MNO. The size of the MNO affects the total surplus to be bargained over, not its division. This is because, once MNOs have some subscribers, bargaining might occur over calls destined to those customers, therefore without substitution possibilities. As a result, we can conclude that the existence of countervailing buyer power over the setting of termination prices does not seem more likely for small MNOs ¹².

■ Conclusions

Practitioners and policy makers should not forget that the role of market definition is to provide a basis on which regulators or anti-trust authorities calculate important indicators such as market shares, etc., in making their prima facie case. However, one should be very careful not to make too much of market delineations. Market definition is not a substitute for a full analysis of the likely competitive effects in a certain economic environment under examination. The task of defining markets should not be confused with the assessment of competitive effects and efficiencies. In practice, this means that many subtle interactions that may be missed when defining markets as a first cut, can be taken into account at later stages, for example, when assessing market power and eventually imposing remedies.

¹² In fact, there are theoretical arguments (and some empirical evidence) for supporting the opposite result: smaller networks charge more for F2M termination than bigger networks in the presence of consumer ignorance, mobile number portability, or no discrimination requirements for F2M calls. See GANS & KING (2000) and WRIGHT (2002).

References

ARMSTRONG M.:

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

- (2006): "Competition in Two-sided markets", *RAND Journal of Economics*.

BINMORE K. & D. HARBORD (2005): "Bargaining over fixed-to-mobile termination rates: countervailing buyer power as a constraint on monopoly power", *Journal of Competition Law & Economics*.

CAMBINI C. & T. VALLETTI (2005): "Information Exchange and Competition in Communications Networks", CEPR Discussion Paper.

EVANS D. (2003): "The Antitrust Economics of Multi-Sided Platform Markets", *Yale Journal on Regulation*.

EVANS D. & M. NOEL (2005): "Defining Antitrust Markets when Firms Operate Two-sided Platforms", *Columbia Business Law Review*.

GANS J. & S. KING (2000): "Mobile Competition, Customer Ignorance, and Fixed-to-mobile Call Prices", *Information Economics & Policy*.

LAFFONT J.J. & J. TIROLE (2000): *Competition in Telecommunications*, MIT Press, Cambridge (MA).

ROCHET J.C. & J. TIROLE (2003): "Platform Competition in Two-sided Markets", *Journal of the European Economic Association*.

VALLETTI T. & G. HOUPIS (2005): "Mobile Termination: What is the 'Right' Charge?", *Journal of Regulatory Economics*.

VOGELSANG I. (2003): "Price Regulation of Access to Telecommunications Networks", *Journal of Economic Literature*.

WRIGHT J.:

- (2002): "Access Pricing under Competition: an Application to Cellular Networks", *Journal of Industrial Economics*.

- (2004): "One-sided Logic in Two-sided Markets", *Review of Network Economics*.