Obligations that can be imposed on operators with significant market power under the new regulatory framework for electronic communications

Access services to public mobile networks

September 2003

Tommaso Valletti
Imperial College London
t.valletti@imperial.ac.uk

1. Introduction

Under the new regulatory framework for electronic communications networks and services, National Regulatory Authorities (NRAs) have to analyse the defined markets in order to determine if they are effectively competitive. A finding of effective competition in a defined market means that there is no dominant firm operating in that market. If the market is effectively competitive, NRAs shall withdraw any existing obligations and not impose any new ones under the SMP process. On the other hand, if the market is not effectively competitive, NRAs will designate the dominant operators as having SMP, and impose on them the appropriate obligations.

In case of wholesale markets, NRAs will choose obligations among the menu provided for in Articles 9-13 of the Access Directive (transparency, non-discrimination, accounting separation, access, price control and cost accounting), or can impose any another obligation with the prior agreement of the Commission.

The purpose of this study is to provide some guidance in the choice of remedies for mobile telephony. The analysis conducted here is driven by economic principles. For each defined market, the nature of the problem is first described, followed by the understanding of the potential market failures associated to that market. We will then discuss the optimal approach, defined as the one that should be enforced by a NRA in the fortunate position of having all the relevant information on cost and demand at its disposal. Finally, various remedies will be assessed against this benchmark. The main focus is on the understanding of the economic effects of the various remedies proposed, including unintended side-effects that may arise from a certain form of intervention. The proposed remedies have both a normative and a practical intent, hence this paper's aim is to highlight their effectiveness and their practicality.

The following three mobile markets defined in the Commission Recommendation are covered in the remainder:

- Wholesale markets of access and call origination on the public mobile networks (Section 2);
- Wholesale voice call termination on individual public mobile networks (Section 3);
- Wholesale international roaming on public mobile networks (Section 4).

2. Wholesale access and call origination on public mobile networks

For the last twenty years, mobile telephony has been, along with the Internet, the technology that has most dramatically changed the telecommunications sector. Mobile communications started as a premium service offering voice transmission with mobility. As the service became more common, mobile telephony challenged the notion of natural monopoly within the sector, and this has unravelled a wave of regulatory change that has deeply changed the market structure of the whole telecommunications industry (Laffont and Tirole, 2000). In many countries there are now more mobile phones than fixed lines.

The key variable for the development of the industry is radio spectrum, a scarce resource that could be used for other purposes. The evolution of the industry is a race for relaxing this constraint. Technological innovation, such as the introduction of cellular technology and subsequent switch from analogue to digital transmission, has helped to make the spectrum constraint less severe (i.e. employing less spectrum per unit of information transmitted). Moreover, technological innovation also helped to extend cellular mobile telecommunications into higher frequency bands of the radio frequency spectrum, which previously were not feasible for providing mobile telecommunications services. This provided more scope for accommodating new users and thus to extend the market size. On the supply side, relaxing the spectrum constraint permitted an increase in the number of firms in the market, with beneficial effects on service quality and prices for the user.

The spectrum constraint has been relaxed, but not yet eliminated, thus entry into the industry is still restricted by regulation and market structure is oligopolistic.[1]

[1] See Hausman (2002) and Gruber and Valletti (2003) for surveys on the economics of mobile telephony.

2.1 The problem

This section discusses the causes of potential problems that may need to be addressed with appropriate remedies in the context of wholesale access and call origination. These are wholesale markets and arise as there is a corresponding retail market for access and call origination.

Access and call origination are analysed jointly as most consumers are interested to buy a bundle of services, made of both access and call origination that, in economic terms, are complements. Also on the supply side they are both offered by MNOs.[2]

Problems can arise only if competition is not effective at the retail level, in which case remedies may be imposed at the wholesale level.

2.2 Sources of market failure

1. Single dominance

This is a standard case of market failure. A firm could try to exploit its market power by setting prices that either 'too' high or 'too' low.

Prices are typically too high when a firm can extract excessive profits from its customers without fearing the loss of market share to its rivals. This may happen when a firm controls most of the market and there are entry barriers of various nature. The most important of such barriers in this particular market is spectrum and licence scarcity. In normal market situations, when prices are too high other firms will be attracted by the super-normal profits earned by incumbents. Alternatively, the simple threat of entry may discipline the incumbents' behaviour not to set excessive prices. This mechanism is somehow broken in mobile telephony as the number of competitors is chosen by NRAs. Outsiders cannot enter this market as a reaction to super-normal profits.

[2] The same considerations apply to text messages: customers care about the price of both components, access and call origination. SMS should then be part of the same market.

Conversely prices may be too low when a dominant firm is engaged in an attempt to predate against a rival. Predation occurs when an incumbent firm charges a low price that is not optimal in the short-run, with the intent of increasing the probability of exit of its rivals. The incumbent renounces to short-run profits in order to obtain increased market power in the future. Also this case has not been particularly relevant in the past, but may become more prominent in the future as new players enter the market. What is important to recall from economic theory is that the analysis of an alleged predatory situation is rather complex and should look at a certain number of fundamental elements:

- There should be structural factors that allow the incumbent firm to exert market power and raise prices in the future;
- There should be a feasible scheme of predation (e.g. the prey has financial problems);
- The should be a probable recoupment of the losses incurred during the predatory phase;
- There should be some cost information (e.g. whether the price is below some measure of cost: this is relevant but cannot be the only piece of evidence in a predatory case).

2. Joint dominance

When firms compete repeatedly against each other, they may recognise that their repeated interaction can be used to sustain high prices in a 'tacit' way, i.e. without any explicit co-ordination. Whether or not tacit collusion is easy to enforce depends on a fundamental trade-off: the 'gain' from deviation against the 'punishment'. To understand this, imagine one firm that is tacitly colluding: this firm sets high (possibly monopoly) prices and shares the corresponding profits with the other firms in the market in every period. If this firm deviates from the agreement to sustain high prices, it can get a good share of the market before the rivals can react (this is the 'gain' from deviation: in the limit it could get almost the

full monopoly profits alone for a while). However when the deviation is detected, the rivals will be able to retaliate and punish the deviant firm by setting competitive prices: hence in the future the deviant firm will have to renounce to its share of collusive profits it could have obtained had it not deviated in the first place.

Whether firms can in fact sustain this type of tacit collusion depends on a series of factors including:

- Entry barriers and number of players: these facilitate tacit collusion as new firms cannot enter in the presence of super normal profits and collusion is easier to sustain with a limited number of firms. Given the spectrum constraint, these facilitating factors are indeed present in mobile telephony;
- Frequency of interactions and transparency: these help sustain collusion as they make the gains from deviation short-lived. In mobile telephony, interactions are very frequent and prices can be adjusted quite rapidly. On the other hand, it is not clear how transparent these tariffs are: they are easy to observe, but is much less clear how customers react to new packages, hence a deviant firm may use this fact in order to postpone as much as possible the rivals' retaliation.
- Multi-market contacts and structural links facilitate collusion as punishment can be made harsher (it can be imposed in many markets). This aspect is becoming increasingly relevant in mobile telephony as a few big groups operate across many European countries.
- Market growth makes collusion easier to sustain as firm do not want to miss the opportunity to share profits in the future, when market size will be bigger. This facilitating factor is probably present for 3G, not for the more mature 2G mobile telephony.
- Symmetry makes it easier to collude. It is easier to understand this by noting that asymmetries will make one of the players more likely to deviate (typically,

the most efficient firm). The relevance of this facilitating factor in mobile telephony can only be assessed on a case by case basis.

3. First mover(s) advantage

This aspect is particularly relevant during the transition phase from 2G to 3G services as those network operators with a 3G licence but no 2G infrastructure may be at a disadvantage against incumbent operators while they roll out their networks. The potential problem that arises here is that, in the absence of appropriate intervention, the entrants may become too weak and competition could not fully develop. While this problem is likely to be found in many markets, it is not clear if the current framework allows NRAs to intervene.

At present, either there is SMP or there is not. If no SMP is found, then no remedy can be imposed. The case of a first-mover advantage falls in a grey area: it is indeed possible that no single or joint dominance is found in the retail market, yet competition is not fully fledged as mobile telephony typically involves competition among only a few firms in a given country. These firms could be able to co-ordinate in the wholesale market, say to deny national roaming to a new entrant. These firms are in a way 'jointly dominant' in the wholesale markets as they mutually agree that they should not create their own competitors.[3] However, if no SMP is found in the retail market, then it is not clear why a remedy could be imposed. This could be desirable from an economic point of view, but not feasible from a legal side.[4]

[3] They are not singly dominant in the wholesale market: if one of the existing rivals provides national roaming to a new entrant, then it is likely that all existing operators would want to do the same as by denying roaming they would not be able to impair the entrant and they would simply renounce to eventual roaming profits.

[4] In the UK, Oftel has decided to use Article 5 of the Access Directive for the purpose of imposing a national roaming condition.

2.3 Optimal approach

The potential market failures described above are quite general, hence also the optimal approach can be only described in general terms as well:

- In case of single dominance (excessive prices), the NRA should try to replicate effective competition;
- In case of joint dominance, the NRA should try to restore proper incentives for operators to compete against each other;
- In case of first mover(s) advantage, the NRA should try ensure a level-playing field.

2.4 Remedies

This section contains a discussion of potential remedies applied to the wholesale access and call origination. If competition in the retail market is found to be effective, no specific remedies are required.

Remedy 1: Structural remedies

The presence of single dominance or of joint dominance in the retail market may be a signal of an inappropriate market structure: super-normal profits are earned by one or more firms, not due to their superior ability, but because other operators cannot flow into this market where entry is restricted by licensing. Rents are created by licence scarcity and they cannot be dissipated by the normal workings of competition. A structural remedy would try to remove this system by allowing a flexible market for spectrum, in order for the market rather than the NRAs to decide on the optimal number of MNOs.

Ideally, the current model of centralised spectrum management should be replaced with decentralised solutions. In the new system, the default rule should grant operators with the highest flexibility, with the regulator monitoring the proper working of competition rather than deciding who does what. This type of

remedy goes beyond the powers allowed by the New Regulatory Framework, although there is a radio spectrum Committee which has a range of harmonisation and liberalisation powers. This remedy should be seriously taken into consideration as the current process of spectrum allocation is a source of many inefficiencies that also concern markets other than mobile telephony.

Remedy 2: Prohibition of discrimination

This is an obligation to apply the same conditions in equivalent circumstances and not to discriminate in favour of the regulated MNO's own divisions. It can be an appropriate remedy in case single dominance is found. This remedy needs to be accompanied by an ancillary remedy such as accounting separation in order to ensure compliance (the internal divisions of a MNO should charge equally whether internally or to another party).

This remedy, in the form of a margin test, can also be imposed in order to avoid predatory cases (however, we argued above that several other elements must also be taken into account in a case of predation).

Remedy 3: Offer call origination agreements

This refers to various forms of obligations introduced to meet reasonable requests for access or interconnection. According to the severity of the problem, remedies may take a light form such as an obligation to negotiate, or a more restrictive one such as the provision of Indirect Access or CPS at cost-oriented prices.

This remedy can be imposed if single dominance is found, in which case it would be particularly effective if cost orientation is also required. Pricing rules such as 'retail-minus' would not be particularly effective in this case as an equally efficient (downstream) entrant would just be able to replicate the same retail price, hence 'excessive' prices would still occur in the retail market. On the other hand, cost orientation would allow competition between the entrant and the incumbent to drive prices down. The downside of cost orientation is that, if wholesale prices are

wrongly calculated, they would have a negative impact on investment of the incumbent. This remedy has been applied extensively in fixed telephony where the historical operator had a dominant position and a similar approach can be followed for mobile telephony if similar conditions arise.

Remedies of this sort may also be justified in case of joint dominance: by allowing for MVNOs, for instance, the number of operators in a market would increase, hence making tacit collusion less stable. Also in this case, though, some cost orientation for access would be needed.

The imposition of this type of remedies would have the positive effect of inducing more efficient pricing structures. Their downside is that they would weaken the incentives to invest in competing infrastructure and may simply induce the entry of resellers.

Remedy 4: National roaming

This remedy is a particular case of the previous class of remedies. Its main purpose is to is help new entrants against established incumbents while they roll out their networks. National roaming is expected to become more relevant with 3G, both because of incumbency advantages for those operators with a 2G infrastructure and corresponding subscriber base, and because it is unlikely that expensive 3G coverage will be extensively used in rural areas, hence 2G and 3G infrastructure will coexist.

Since national roaming can be mandated as a way to overcome an initial asymmetric position between players, it is also important not to abuse with its use. This remedy should then be temporary, otherwise there would be reduced incentives to invest for the entrants. As far as its pricing is concerned, this could be done leaving its setting to the parties, using 'retail minus' charges as a dispute resolution.

While a national roaming remedy can be applied when SMP is found, its role is more general and it is appropriate also when there is no SMP, yet

competition is not fully effective. However, in this last case it is not clear that any remedy can be imposed, as no SMP is found in the retail market. However, NRAs are able to impose remedies outside of the SMP process, e.g. Article 5 of the Access Directive could be used instead to impose national roaming if there is no SMP but competition is not fully effective.

3. Wholesale voice call termination on individual mobile networks

The service considered in this section is wholesale mobile voice call termination. This is a network access service needed in order for callers belonging to different networks to speak to each other. If wholesale call termination were not available, operators could then only complete calls to other customers on the same network. Call termination is a wholesale service since it is typically purchased by network operators, and not directly by users.

In particular, there are two types of call termination that will be considered in this section:

- Termination of calls from a fixed-line network to a mobile network (a fixed-to-mobile or F2M call);
- Termination of calls from a mobile operator to another mobile operator (a mobile-to-mobile or M2M call; these are also sometimes referred to as M2M 'off-net' calls to distinguish them from termination of calls initiated and terminated on the same network, M2M 'on-net' calls).

3.1 The problem

Wholesale mobile call termination on an individual mobile network is a relevant market. In such a market each mobile network has the ability to exert market power.

1. Substitutability from the demand side and CPP

As the termination charge typically feeds into the retail price paid by the caller, it is appropriate to consider how final users react to higher prices of call to mobile networks. We need to distinguish between two types of users, those who make calls and those who receive them.

Let us consider the callers first. 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 his/her demand elasticity, but it is very unlikely that the caller can find good alternative substitutes.

To give a few examples, in place of a F2M call, the caller could send a text message instead. However, this would be feasible only if the calling party also owns a mobile phone (or a sophisticated fixed line handset capable of sending SMS). In addition, a text message is short and there cannot be simultaneity in the conversation. While this practice is feasible in some circumstances, it is clear that this type of substitution cannot be generalised. 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.

Calls using Voice over Internet Protocol (VoIP) could also be a substitute for mobile call termination. These calls are treated as Internet data. The sender only pays to be on-line and no termination charge as such can be levied. However, for this to happen, the receiver needs an advanced handset (e.g. GPRS-enabled). More importantly, a VoIP call can be terminated only if both the sender and the receiver are on-line simultaneously, which is unlikely since the receiver is also charged for the time he/she is on-line. Finally, if the receiver has the option to receive a call using either VoIP or with the more common technology, he/she would probably opt for the latter since he/she would then not be for charged it.

Constraints on increases in the price of termination charges could also arise if receivers themselves react to an increase in the price of a call to a mobile. This is very unlikely to happen since a crucial aspect in the mobile industry is that the party making and paying for the call is 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. The mobile owner cares most about the prices he/she has to pay to subscribe to and place calls with a mobile operator, but in most cases he/she will not take into account the prices paid by other callers to contact him/her. CPP is the biggest source of distortion in this market, since it is the ‘wrong’ party that gets the bill: the caller pays, but the mobile network

subscription is decided by somebody else. This creates an externality that leads to market failures that are discussed below.

There are some circumstances in which the receiving party will care about termination rates. For instance, if the receiving party is happy about receiving calls, his/her satisfaction is increased the lower the termination charges are (since they will feed into lower retail prices of calls to mobile, hence more calls will be placed and received). Also, if the receiver cares about the satisfaction of the caller, then the price of calls to mobile 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 sort of telephony services to their employees. These can constitute a large part of the customer base of MNOs, however MNOs have the ability to price discriminate among different groups, for instance by offering discounts to large business users, hence their presence does not constraint overall price levels to other customers. In general, the ability of an operator to set termination charges is not constrained by the behaviour of the called party.

Finally, since wholesale call termination on a certain mobile network cannot be substituted with wholesale call termination on a different network (in which case the call will not be terminated), this type of substitutability does not exist.

To conclude:

- There is limited demand substitution both at the retail and at the wholesale level.
- Demand elasticity can be increased if:
  a) There are closed user groups;
  b) The receiving party cares about being called;
  c) Call-back, or text messages, or M2M calls, can be substitutes for calls to mobile users.

2. Substitutability from the supply side

This type of supply substitutability is also very limited since, in most cases, the originating operator cannot buy call termination on a given network from an alternative source. In principle, wholesale supply-side substitution could happen if access to the SIM card of a customer were available to more than one network. This does not seem to be available at present.

Wholesale substitutability from the supply side could also happen in the presence of a MVNO who offered termination services and could select among alternative mobile networks on which to terminate calls. However, even if this were technologically feasible, this would not change the underlying economic problem: the mobile customer would, also in this case, care most about the prices he/she pays and not about the price paid by other customers to call him/her. The presence of MVNOs could shift termination rents from MNOs to MVNOs but would not eliminate such rents.

3. F2M calls and M2M calls: similarities and differences

The termination of a call to a mobile user initiated by a fixed user (F2M) presents many points in common with the termination of a call initiated by a mobile user (M2M). Most importantly, the CPP feature applies in both cases. SMP (single dominance) is present over termination as the recipient mobile network has market power over the calls it receives from other customers, be they fixed or mobile users.

However, there is one distinction to be made. Fixed and mobile telephony are largely distinct markets, so termination rates will not have any impact on competition between the fixed originating network and the mobile terminating network. On the other hand, as far as M2M is concerned, mobile networks compete against each other for customers. As a result, in the M2M context there are additional aspects that are not present in the F2M case. Since competing mobile operators have mutual interconnection needs to terminate calls destined to each other's network, this means that termination charges can provide one way in which competitors interact, creating scope for retaliation and collusion.

Termination charges have a direct impact on the operators' perceived costs. Hence, they can be used by incumbents to protect their advantages against new entrants. In addition, termination charges can be designed to soften competition.

To conclude, while single dominance is a potential problem both for F2M calls and for M2M calls, joint dominance is a potential problem peculiar only to M2M calls.[5]

3.2 Sources of market failure

1. Fixed to mobile calls: the fundamental distortion

The main source of market failure is that each mobile network operator has monopoly power over the termination of calls on its network. This means that it can set termination charges at the monopoly level without fearing competition from rivals.

This stark result is a direct outcome of CPP pricing arrangements. Once a person has decided to join a particular mobile operator, that operator has a monopoly position over termination services to its subscribers. These services are paid for by other users, not by the subscribers themselves. It is then clear that the decision to subscribe to a network has an influence on the price charged to all other customers that may want to call that person. Hence termination services involve an externality problem that is a potential source of distortions.

This distortion implies that the mobile operator is typically able to set termination charges at the monopoly (e.g. profit maximising) level, independently of the intensity of competition in the market for subscribers. The price of termination is determined by the same trade-off made by a monopoly firm: by setting higher prices it increases the unit margin it can earn, but it also reduces the quantities supplied. The profit maximising charge is the one that solves this trade-

[5] Notice that the single dominance problem for M2M calls can arise in the market for call termination. On the other hand, joint dominance can be a problem in the broader mobile retail market, in which case M2M termination rates can be the instrument used to sustain high retail prices.

off and maximises profits. Termination charges involve the same monopoly trade-off: the mobile operator will increase termination charges until they optimise their profits over calls received (this trade-off is dictated by the demand elasticity of the calling parties).

It is important to recognise that this distortion arises independently of the way MNOs compete against each other. If they do not compete against each other, then all termination profits will stay with the MNO. On the other hand, if there is competition among MNOs, then the termination profits would be passed on to mobile users, for instance via lower rental fees or via cheaper handsets, and the excess profits are competed away. These two different scenarios matter for distributional effects. However, the fundamental distortion remains, no matter how intense competition for customers is. In both cases fixed users are charged inefficiently high prices. In other words, even if there is competition for mobile customers, there is no competition over calls destined to those customers.

Since this is the most important point in the call termination problem, it is helpful to take a stylised situation where the mobile sector is assumed to be perfectly competitive (i.e., operators do not make any super-normal profit) and mobile operators charge two-part tariffs to customers with identical preferences (for instance a monthly fee and a charge per minute for every call made). Also assume, for the sake of simplicity, that mobile users only call fixed users and receive calls only from them. Then operators would compete to attract customers by setting each call origination charge equal to its marginal cost and set the monthly fee to divide the surplus created between the operator and its customers. These results ensue since there is no reason for mobile operators to set outgoing call charges above cost: marginal cost pricing is efficient and firms have another instrument (the rental fee) to eventually extract profits, so they do need impose any distortion on outgoing charges.

If, as assumed, the mobile industry is perfectly competitive, operators would earn zero extra-profits. Any increase in termination profits (for instance because the termination charge is set above its cost) would simply be passed to

mobile subscribers via lower fixed charges. Fixed charges may even become negative, as long as considerable extra-profits arise from termination: this may explain handset subsidies, a common feature in many mobile markets.

However, even if there are no extra profits in equilibrium, each firm will have a unilateral incentive to set the termination charges of calls it receives from fixed users at the monopoly level. In fact, mobile operators will want to maximise termination profits so as to subsidise their mobile subscribers as much as possible. If one operator did not set them at the monopoly level (but the rival did), it will be at a disadvantage and mobile customers will all go to the rival since the latter could offer a better deal, passing them termination revenues, for instance in the form of cheaper handset or rental charges which the first operator had forgone through the lower termination charge.

Some competitive pressure arises, and potential market failures associated with termination services are diluted, if people care about receiving calls. In fact, if a mobile user places similar weights to calls made and received, then any attempt on the part of a mobile operator to set high termination charges would induce subscribers to change network, since they would otherwise receive too few calls. This result is true even if the caller and the receiver do not belong to the same 'closed user group'. The argument that people may care about receiving calls is realistic in the mobile sector since a mobile phone gives a customer the ability to be reached by other people at any time in any place. However, the evidence of this type of behaviour is limited and mixed.

We can thus conclude that even with perfect competition for mobile users, there is little competition for providing access to mobile subscribers. This remark suggests that if mobile operators are free to determine termination rates, they will set charges that extract all possible surplus from fixed users. Thus, even when there are no extra rents overall, the wrong price structure would arise.

While we reiterate that the basic distortion, due to CPP, remains no matter what kind of competition prevails in the market, the optimal approach cannot be always determined irrespective of such competition. In other words, while the

basic distortion exists in the market for call termination on an individual network, its assessment must include also other markets, notably the whole retail market for mobile services. This is not because the retail market constraints in any way termination rates, but because the effects of the distortion in one market (e.g. call termination) may have an impact also in another market (e.g. call origination).

2. Fixed to mobile calls: a second distortion

A second source of distortions is related to consumer ignorance. Fixed-line users may have little knowledge of the mobile network they were calling and of the call price. If fixed-line users base their calling decisions only on an estimated price based on mobile market shares, then the link between a specific termination charge set by a network and the number of calls terminated on that network is broken. If a mobile network raises its termination charge, it gets the full benefit and shares with other mobile networks the reduction in the number of calls received. As a consequence, networks will have an incentive to set very high termination rates, even above the monopoly level. In fact, as termination charges are increased above the monopoly level, two additional effects follow. On the one hand, termination profits per subscriber will decrease, which has a negative impact on the mobile operator's profit. On the other hand, the increase in termination charges will increase the F2M price of all the calls, which will also decrease the termination profits per subscriber which rivals can capture. This causes rivals to compete less aggressively for mobile subscribers. When both operators set termination charges at the monopoly level, the first effect is zero, while the second effect remains, implying that each operator will want to set termination charges above the monopoly level.

Notice that this problem is formally equivalent to an environment where a uniform fixed-to-mobile price is set since the increase in one particular termination charge feeds into an increase of the price of calls to all mobile networks. Hence this second source of distortion creates problems in environments where it is difficult for a caller to identify the network he/she is calling. This can arise for various reasons:

- consumer ignorance;
- mobile number portability;
- no discrimination requirements for F2M calls.

3. Mobile to mobile calls

The call termination problem described in the previous sections is rather extreme in the sense that it is relevant when the market of callers is completely separate from the market of receivers (e.g., mobile users and fixed subscribers). In the case of M2M calls, operators compete for the same customer base that both originates and terminates calls. As long as operators *A* and *B* both command some market share, operator *A* needs interconnection with *B* to terminate the calls that *A*'s customers destine to *B*'s customers and vice versa. There is a sort of "double coincidence of wants" that potentially makes the interconnection problem less problematic. In a symmetric situation termination charges may even be thought to be irrelevant since *A* pays *B* the same amount it receives from *B*. However, this reasoning is not entirely correct.

First of all, we need to distinguish between termination charges set unilaterally or jointly. The latter could arise, for instance, under a reciprocity clause. In the former case, there may be a typical double mark up problem. However firms may negotiate more sophisticated contracts to get rid of this inefficiency.

When termination charges are set jointly, two kinds of problems emerge:

- Operators can agree to set access charges at a level that eliminates any effective competition among them;
- Termination-based discrimination creates forms of externalities that may be used especially against small new entrants.

The first concern arises when termination charges can be used as an instrument of tacit collusion. Collusive (i.e., monopoly) prices can be sustained using high termination charges because of a "raise-each-other's cost" effect. To see this, imagine what happens when operators charge monopoly retail prices to customers. If mobile customers call each other with the same probability, the traffic is balanced and an operator pays the rival for termination services the same amount 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 are destined to 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. To get this result some conditions are needed, for instance products need to be not too homogeneous, otherwise the incentive to undercut would have the additional benefit to get market share.

Perhaps more crucially, another condition that is needed to generate this non-cooperative collusive result is that retail tariffs are linear. Once firms are allowed to offer non-linear prices the result collapses. For instance, with two-part pricing, it is still true that a high termination charge feeds into high retail charges. However, all the profits generated are used to lower the fixed component. Pricing itself may become efficient, since operators would tend to charge call prices equal to their perceived marginal costs: this result typically occurs when operators have more instruments to build market shares without having to inflate their outflow charges.

As far as (reciprocal) M2M calls are concerned, we then have a double set of results. On the one hand, these reciprocal deals give an instrument to firms to influence competition at the retail level. This instrument can be abused to tacitly collude in the retail market (recall that this result is not very robust to sophisticated retail pricing structures). On the other hand, there is also a mutual interest to set

these charges at efficient levels as firms recognise that excessive (reciprocal) termination charges will leave some gains unexploited. The degree of symmetry between networks will also crucially affect the incentives to set one particular type of (reciprocal) termination charges, since a net receiver of traffic will likely prefer above-cost termination charges and viceversa.

The second concern arises when mobile operators price discriminate between calls destined on-net and off-net. While it is well-known that price discrimination may be dictated by efficiency reasons reflecting customer heterogeneity, their downside is that they can also be used anti-competitively. For instance, if consumers care about incoming calls, off-net charges would tend to be set at high levels in order not to give rival's subscribers benefits from receiving calls. High off-net termination charges make the rival network less attractive to subscribe to. High access charges could even cause a *de facto* connectivity breakdown.

We conclude by noting that both termination "collusive" charges and termination "discriminatory" charges are problems which bring inefficiencies in the retail market, but which may need remedy in the wholesale market.[6]

3.3 Optimal approach

Welfare considerations on termination rates are quite complex, since an increase in termination charges both increases the price of fixed-to-mobile calls and may decrease fixed fees and other retail charges that apply to mobile users.[7]

---

[6] Notice that so far we have treated M2M calls as traditional two-way interconnection problems. However, with the development of mobile telephony, M2M calls can become a substitute for F2M calls. If customers are charged excessive prices for making F2M calls they could opt for M2M calls instead. Under this scenario, this substitution possibility would constrain the ability of MNOs to charge high F2M termination charges to the extent that M2M calls are in line with their costs. Anticipating this kind of customer arbitrage, MNOs may want to oppose it by setting high M2M termination charges, more or less to the level of F2M termination charges. Hence off-net price discrimination may emerge for reasons other than the "collusive" or the "discriminatory" effects described above.

[7] This problem has been analysed quite extensively in the economics literature (Armstrong, 2002; mason and Valletti, 2001; Wright, 2003).

The main inefficiency that arises is that F2M calls are typically charged too much so that a suboptimal allocation of F2M calls occurs. If this is the only inefficiency, the socially optimal F2M termination charges are then equal to cost. This ensures that also F2M calls are priced at cost. Marginal cost pricing (implying no subsidies for mobile connections) is the correct benchmark when a certain number of assumptions are satisfied:

a) the demand of mobile subscribers is rigid with respect to subscription decisions (i.e. the mobile market is 'mature' and always fully 'covered') and there are no network externalities;

b) there is no monopoly power exercised by the fixed network;

c) there are no call externalities; and

d) there are no fixed costs to recover.

We now consider how the socially-optimal benchmark changes when the previous conditions are not satisfied, in which case there are departures from cost-based termination charges.

When condition a) is violated, then above-cost charges would be beneficial in the presence of network externalities, since higher termination revenues could be used to subsidise entry, thereby raising the equilibrium number of subscribers - this benefits everybody, because both fixed-line and mobile customers are able to call and be called by additional subscribers. As mobile subscription tends toward saturation point, the need to effectively "tax" F2M calls to subsidise mobile subscription diminishes.[8]

If condition b) is not satisfied, then F2M retail calls already involve a margin above the termination cost, as the fixed-line network is able to exercise

[8] A well-known result in taxation theory says that efficient mark-ups should decrease with an increase in the tax-base.

some market power. These mark-ups over termination charges added by fixed operators should then be offset by setting termination charges below cost.[9]

If condition c) is not satisfied, then optimal termination charges are lowered with respect to the benchmark, since then mobile subscribers benefit from lower F2M prices. Notice that in this scenario MNOs themselves will have a lower incentive to set high termination charges since consumers would be less willing to join a network that offers cheap subscription but makes it expensive for other callers to call the subscriber.[10]

The benchmark ignores any role for the fixed and common costs of operating a mobile network (condition d)). Implicitly, it assumes that retail mark-ups are sufficient for recovering fixed and common costs. In the face of these costs, optimal Ramsey termination charges become the correct benchmark and they typically include an extra mark-up above termination costs, which would enable prices of other services to be lowered.[11]

It is useful to clarify some points related to Ramsey charges that have appeared in recent court cases:

- Ramsey charges mean that fixed and common costs[12] should be recovered according to the "inverse elasticity rule", i.e. higher mark-ups are imposed

[9] This result basically says that it may be optimal in some case to "subsidise" a monopolist (the fixed-line network in this case). This type of subsidisation ceases to exist if the original market power is addressed directly in the relevant retail market (fixed call origination in this case). Similarly, this "subsidising a monopolist" approach is also applicable to MNOs to the extent that competition among them is imperfect, in which case higher termination charges would be called for.

[10] If both the caller and the receiver benefit from a call, they should both contribute to its payment. Under a CPP system this does not happen, hence 'too few' calls are typically placed since the caller does not internalise the benefit of the receiver. DeGraba (2003) shows how a 'bill-and-keep' system would lead firms to reduce the price of outgoing charges, and recover the termination costs directly from own customers, thus restoring efficiency.

[11] No Ramsey mark-up is required when demand for mobile subscription is totally inelastic in which case all fixed and common costs should be recovered via monthly rental fees, without inducing any additional distortion.

[12] The calculation of fixed and common costs is a very difficult task in itself. If a cost can be directly attributable to a service, then it should be recovered directly via the price of that service. A case in point is customer acquisition costs: if they are truly fixed costs, then they should be recovered via various mark ups on the different services. However, if they are caused by an endogenous choice of the network operator, they should not be attributed to the pool of common and fixed costs, in order to avoid a circularity problem: the current level of customer acquisition

in those segments where demand is less sensitive, so that resulting allocations are not particularly distorted. These tariffs are sometimes referred to as "2nd best", since the "1st best", i.e. marginal cost pricing, would not allow to break even and balance the budget.

- Firms also typically set their optimal (i.e. profit maximising) mark-ups according to the inverse of the elasticity of the demand they face. However the firms' elasticities and the industry elasticities do not coincide. Hence, even if a firm privately sets prices with a Ramsey structure, this does not imply that it also sets them at the socially-optimal Ramsey level.
- Ramsey considerations are at least as important for the fixed-line operator as they are for mobile operators, hence the Ramsey argument alone does not provide an automatic justification for high F2M termination charges.
- The welfare-maximising Ramsey charges may take different values according to the number of instruments that the social planner has at its disposal. In one scenario, the social planner may be able to optimise over all prices, including outgoing mobile charges and monthly rental fees ("2nd best"). In a different scenario, the retail charges are set by the MNOs, while the social planner may control only termination charges ("3rd best"). This latter scenario is typically the most relevant. In both cases, though, it is possible to talk about Ramsey charges, since elasticities are taken into account in welfare maximisation. There is no unequivocal relationship between 2nd and 3rd best Ramsey charges, and their levels depend on the parameter values used in the calibration exercise.

We conclude with a brief discussion of optimal M2M 'off-net' termination charges. Cost-based termination is the efficient benchmark in most cases, as efficiency requires not to distort calling patterns once subscription decisions have

---

cost may in fact depend on the level of termination charges. In other words customer acquisition costs should not be included in the Ramsey calculation if it is likely that such costs arise because of

occurred. In the presence of fixed or common costs, a mark-up above costs may be needed under some circumstances (this is the same argument developed above with respect to Ramsey "2nd best" charges). However, there are no particular reasons to believe that on-net and off-net (retail) prices should differ for efficiency reasons since the underlying elasticities would be very similar for both types of calls.

3.4 Remedies and their consequences

This section contains a discussion of potential remedies applied to the wholesale call termination market.

Remedy 1: Cost-based regulation (cost recovery, including charge controls)

Termination charges based on marginal costs (LRIC) are an option that, as argued before, can be legitimate in the absence of high fixed costs. Charge controls are appropriate to offset the lack of competition in the market for call termination on individual networks.

This remedy would solve the market failure since it would be possible to directly regulate termination charges and the price of calls to mobile users would be lowered to efficient levels. In general, this remedy would have a positive first-order benefit on the callers, but may have a negative second-order impact on receivers. Overall, welfare would increase.

The downside of this kind of intervention is that it is typically rather heavy handed. It can also be put in place only if cost-proxy models for mobile networks are available and if demand elasticities are known.

A related type of charge control that puts a downward pressure on termination charges is a price cap mechanism. There are many options within this category

---

the high termination charges. As we discussed above, this is one of the mechanisms by which excess termination profits are competed away in the retail market.

since the basket of services subject to price control may include many separate caps for termination services, one single cap for all termination services, a cap for all wholesale services, or a cap for both retail and wholesale services (i.e., a global price cap[13]).

Remedy 2: Yardstick

The Australian regulator ACCC decided in 2000 that any discount that mobile operators offer to their customers would have to be passed also on termination. The advantage of this remedy is that the benefits from competition with respect to subscription and call origination would be passed on to users that place calls to mobile. However, the downside to this is that, in the anticipation of the additional effect, operators would be more reluctant to fight against each other to attract mobile customers. Hence this remedy risks distorting the competitive aspect of mobile telephony as it would expand regulation also into areas not previously covered.

A second and probably 'better' yardstick is to tie termination to the average of on-net calls of other operators. Also in this case, however, there is a need to estimate measures of termination costs to calibrate the optimal initial levels and weights applied to the tying mechanism. In this respect, the apparent advantage relative to cost-based regulation partly disappears.

Remedy 3: Receiver Party pays

Since the cause of the market failure in this market is CPP, and since the fixed-to-mobile termination problem has not arisen in North America under RPP, this is an alternative pricing arrangement that regulators might want to consider. This remedy could solve the problem as receivers, by being charged, would become sensitive to termination charges and prices would go down.

[13] A global price cap is the preferred solution from the theoretical point of view, although there is very little practical experience with its application.

A downside of this remedy is that growth rates have been slower in North America compared to Europe. However, given the current level of penetration, this makes RPP arguably an interesting option for 2G mobile telephony since there is no further need to subsidise the subscriber base.

There is a second downside to this remedy that may produce unintended effects. Recall that the intervention is lead by the desire to pass some more benefits on to fixed subscribers. Under RPP, it may happen that - despite the decrease in the price of fixed-to-mobile calls would increase attempted calls - the actual number of completed calls is diminished since the receivers would keep their handset switched off more often than under CPP. Hence fixed users may be worse off under RPP. RPP solves the termination problem as it results in lower charges but may produce sub-optimal usage of mobile phones. This remedy has to be assessed with respect to the behaviour of both the called and the calling parties.

Remedy 4: Technological solutions: multiple SIM cards

The call termination problem exists because operators do not compete directly over these types of services. Some pressure on the level of termination charges could be placed by giving the customer the opportunity to choose two operators - one for origination and one for termination of calls. This could happen if the called party is able to switch his/her handset between different networks, for instance by having multiple SIM cards. This remedy removes the bottleneck and creates direct competition in both markets. This remedy is also much less intrusive than direct price controls.

The remedy could work since, under CPP, the customer would still be financially responsible only for outgoing calls and hence will choose the operator that offers the cheapest package for originating calls close to the customer's profile. The customer will also be inclined to choose separately the cheapest package offered by competing operators to terminate calls since he will anticipate that more people will be willing to call him/her.

There are two downsides to this remedy. First, there would be some costs associated with this unbundling proposal, for instance the handset would have to contain two SIM cards. This is a software problem that may have some technological difficulties. Second, one operator could still offer both components of the bundle, making it more attractive for a customer to buy the whole bundle from the same operator. Some controls on retail price discrimination and bundling would then be needed in order to prevent MNOs from neutralising this remedy through retail pricing strategies.

A related remedy is to allow MVNOs to provide termination services to subscribers. This would put pressures on MNOs to offer MVNOs lower termination charges. However, this remedy is not likely to work (unless unbundling is required as well) since once customers subscribe to a MVNO, then the standard termination problem arises once again.

Remedy 5: Technological solutions: arbitrageurs

It is possible to convert calls to mobiles onto on-net mobile calls by using particular gateways. This remedy could work when there are arbitrage opportunities generated by low on-net prices. Similar mechanisms could also arise using various call back services.

This remedy is inefficient since gateways use additional resources to complete the call. However, the simple availability of this remedy could be used to discipline the behaviour of the MNOs.

Remedy 6: Bilateral and reciprocal agreements

This remedy consists in the obligation for operators to engage in bilateral negotiations to set identical reciprocal termination charges. This remedy could be applied to M2M 'off-net' calls. It is not appropriate for F2M calls since the environments for the two sets of charges are very different.

This remedy would work to remove double mark-ups that typically occur when charges are set unilaterally rather than jointly. By setting reciprocal charges, one operator could not gain any advantage over the rival by increasing termination charges. In fact, as gains from trade are maximised when calls are priced at marginal costs, it is possible that operators jointly agree on efficient termination charges.

This remedy has some downsides. Firstly, there is a potential concern for collusion among incumbents, in which case termination charges could be used as the instrument to alter the intensity of competition in the market. Secondly, there could be a concern for discrimination against entrants using high off-net termination charges. This last case could be addressed by also imposing a margin squeeze test. Thirdly, it is not clear how MNOs could reach an agreement when they are asymmetric and the traffic between them is unbalanced.

Remedy 7: Price transparency

This remedy promotes carrier identification. As such, it does not address the fundamental distortion in call termination. On the other hand, it may be used to address the second distortion related to consumer ignorance about the called network, something that is particularly realistic with mobile number portability. We recall that, when there is no correspondence between a certain number and the current network choice of a subscriber, the call termination monopoly problem is made even worse.

This remedy may be practically put in place by using some call identification mechanism, e.g. a voice message or a sound that identifies a particular network or a real time cell display of charges. The voice message could be limited only to those numbers that have been ported. Alternatively, there may be free numbers that could be called/texted in order to recognise the current network identity of the called number.

Remedy 8: Voice mail

This remedy consists in allowing the calling party to leave a voice-message for the mobile subscriber instead of communicating to him/her directly. For instance, this could be achieved with the introduction of voicemail access at certain prices (usually the fixed-line rates). This remedy could also work as a call-back service as the called party could then return the call. The rationale for this remedy is to increase the substitution possibilities for call termination on mobile networks. This remedy could work in closed-user groups. However we argued above that in this case the problem would not be too relevant to start with. In general, this remedy is not likely to be very effective since it does not remove the bottleneck nature of termination. It should also be mentioned these services could already arise at present for purely commercial reasons. The fact that they have not made a lot of ground in practice signals that, even if adopted, they would not change the current scenario in any substantive way.

Remedy 9: Bill-and-keep

This remedy consists in providing reciprocal termination charges free of charge. The justification for this remedy includes some obvious points such as the savings of transaction and measurement costs. In this respect, bill-and-keep does not require spending any money on building cost-proxy models, estimating demand parameters, etc., hence leading to a desirable reduced role for regulatory intervention. 'Bill-and-keep' has also more powerful economic justifications in the presence of call externalities: those networks that terminate calls would be forced to cover the termination costs by charging the called party directly. They could do so by increasing the subscription fees, or by billing called parties directly, introducing RPP billing systems without NRAs having to mandate them.

A 'bill-and-keep' remedy does address the call externality problem but may fail to address the eventual subscription externality problem; thus it seems more appropriate in mature networks. It is also fair to say that there is no claim that 'bill-and-keep' is an efficient solution, especially if networks are very asymmetric, but

neither would a more intrusive remedy such as charge controls in the absence of detailed information about costs and demand structure. A disadvantage of 'bill-and-keep' is that network providers would then try to hand over a call to the other network as soon as possible. In order to avoid this free riding problem, the practical proposal of a 'bill-and-keep' system restricts it to the termination of calls from the last central office to the receiver, while the originating network would be responsible for transport and switching until that point (DeGraba, 2000). Also notice that 'bill-and-keep' has been advocated as an interconnection regime applied to all telecommunications services, both fixed and mobile.

3.5 Specific questions

Some specific questions have repeatedly arisen in this particular context.

1. Text messages

The economic problem of sending and receiving a SMS is similar to voice. It would then seem appropriate to treat them in the same way. In particular, if a difference is made between M2M and F2M calls, SMS are much closer to the former than to the latter.

2. 2G versus 3G

Also in this case the economic problems associated are exactly the same. Hence they have the same economic analysis and the same form of solution. Items which may differ, however, are the cost and the demand parameters. Hence, even if the form of the optimal solution is the same, the actual level of optimal charges may be considerably different.

For instance, subscription elasticity is likely to be quite high for 3G, while it is lower for the more mature 2G. In this respect, the termination mark up should be higher for 3G in the presence of network externalities. In addition, retail substitution may increase for 3G.

It is unclear how 3G termination will evolve, and the problem may be made much less relevant by having increased possibilities of substitution. There is also a lot of uncertainty associated with 3G. These are new services, and their cost and demand parameters are not yet known. Hence the impact on investments cannot be neglected as a 'wrong' regulation risks to affect in an adverse way an infant market.[14]

These arguments then lead to a leaner approach for 3G termination, at least until better information becomes available. However, it is also important to realise that no 3G regulation will also create regulatory loops: operators will be able to set high unregulated 3G termination charges, use extra profits to attract old 2G customers on new 3G services, and avoid being regulated in 2G. The absence of regulation on 3G can be used to bypass eventual regulation on 2G, and 3G subscription would happen 'too soon' as a regulatory artefact.

[14] However, the infant industry argument cannot be stretched too far by allowing, for instance, a mark up on 2G termination to subsidise 3G investments. 3G should stand on its own merits.

4. Wholesale national market for international roaming on public mobile networks

4.1 The problem

International roaming is a wholesale service that satisfies a derived demand by mobile network operators. Demand is derived as MNOs want to satisfy their customers' demand and to ensure them the ability to place and receive calls when in a foreign country.

In principle, customers can find alternative ways to make and receive calls when abroad. For instance, they could use fixed telephones abroad; international prepaid cards and call-back services are quite popular for this purpose. Alternatively, they could subscribe to a foreign mobile telephone service. However the substitutability from both these sources is quite limited. The former (fixed lines) do not guarantee the ability to be contacted anywhere/anytime associated with a mobile phone. The latter (foreign mobile contracts) imply that the customer would need a new telephone number whilst in the foreign country, hence making it difficult for other people to contact him/her. This alternative is also impractical (and expensive) if one travels abroad only occasionally.

There are some entry barriers in this market as only nationally licensed operators can compete to provide roaming services. Indirect access through carrier selection is not a supply substitute at present.

There are two aspects that are quite peculiar to this market, and central to understanding its development - one is institutional and one is technological.[15]

1. The GSM Association.

Most international roaming agreements are concluded on a bilateral basis between individual licensed mobile network operators who are members of the GSM Association (GSMA). The GSMA provides a framework for roaming agreements

[15] See European Commission (2002) and Salsas and Koboldt (2002) for details.

and facilitates negotiations.[16] In 1996 the GSMA introduced the Standard Terms for International Roaming Agreement (STIRA). In particular, the STIRA lays out the tariffing principles for the setting of inter-operator tariffs (IOTs). As a result, some general features have emerged among the participating countries:

- Agreements are usually reciprocal;
- Roaming agreements may include a 'preferred roaming' status;[17]
- There is a non-discrimination clause: each operator applies the same terms and conditions in its international roaming agreements to provide roaming to its network;[18]
- IOTs are published by the GSMA, they are very transparent and available to all operators;
- The ultimate result of this system is that IOTs tend to be very similar across operators within a given country.

2. Technology

On the demand side, consumers are generally poorly informed about roaming conditions and retail prices. Manual choice of a network, while a possibility, seems to be of negligible importance in practice, and it is not clear that, even if both aspects were sorted out, the resulting picture would change. This is because, given the current state of technology, it is very difficult to guarantee traffic re-direction. Network selection is typically driven by the intensity of the signal and/or by the principle of 'last network prevalence'. This results on a selection that is largely random and has little bearing on the roaming charges paid by the end user.

The nature of the problem is best understood by keeping the previous two aspects in mind. Whether or not a problem arises in this market depends on

[16] The framework provided by GSMA does not deal with international roaming agreements between MVNOs and Service Providers. To date no such agreements are in place.
[17] SIM cards can be programmed in such a way as to direct customers to the preferred roaming partner's network (but customers can still manually choose another network operator).
[18] Non-discrimination is a more general requirement imposed by the GSMA on all its members.

whether there are proper incentives for mobile operators to negotiate lower wholesale roaming charges. Incentives arise for visiting networks to look for better deals if they can get a competitive advantage over their national competitors. Incentives also arise for visited networks to offer better deals only if they can generate more traffic. The interaction between the non-discrimination clause (so that if one operator makes a better deal the same terms must be offered to its rivals as well) and the technological inability to guarantee traffic re-direction results in a market that is quite frozen, where no one has a real interest in changing the current pricing structure.

4.2 Sources of market failure

Failures can arise in this market as firms do not have an incentive to respond to the presence of mark ups by reducing prices until they reach their efficient level. Even if competition for customers is present, it is not likely to be effective with respect to wholesale roaming charges. This is due to the two main reasons highlighted above: those who seek international roaming have no big incentive to look for cheaper charges, and those who offer international roaming have no big incentive to offer cheaper charges. Let us consider each assertion in turn.

Under the 'no discrimination' clause there is no major incentive for a visiting network operator to look for cheaper wholesale prices. Of course, cheaper roaming has a direct benefit on the firm’s profit. However, in a competitive context, cheaper costs are usually pursued in order to get an edge over rival operators (as the firm in question will be able to offer cheaper retail charges) and capture a good share of the customer base. This ‘healthy’ mechanism breaks down in international roaming: any discount will also accrue to all visiting network operators from the same country. Hence the firm that put some effort into bargaining a cheaper roaming agreement will get no competitive advantage over the national rivals, since these rivals will also be able to get the same roaming conditions. In the end, the benefits will be the same for both own and rival

customers, hence the deviant firm will have no particular reason to put resources into finding better deals.[19]

As far as the visited network is concerned, the inability to direct traffic to the cheapest of several networks in a visited country is a potential source of market failure since quantities may not follow cheaper prices. If demand did respond to cheaper prices, then a visited operator would want to offer good roaming deals to visiting networks by undercutting rival operators as this would attract additional traffic. However, if demand does not or cannot respond to cheaper prices, then this standard mechanism will not work.

Notice that demand does not respond to prices if technology does not work to re-direct traffic or if customers are poorly informed about prices. Customer poor information reinforces the negative aspects of both market failures explained above. In the first case, the visiting operator could not transmit to its customers a better deal, even if available. In the second case, the visited network would not attract traffic by offering better deals, even if technology did work.

A third potential source of concern in this market is joint dominance. To the extent that a roaming customer perceives no differentiation between foreign networks (which is likely), then all visited networks in a given country can be seen as a collective entity by both the home network and the consumer. Such a collective entity then has a dominant position in the wholesale market. In addition, there are multi-market contacts via the GSMA and there is information sharing as IOTs are highly transparent. Both aspects facilitate tacit collusion. Many contractual clauses, such as reciprocity, also offer a system of retaliation that can be used to sustain high wholesale roaming charges.

A final source of complication arises from the behaviour of NRAs. Since welfare effects are across boundaries, regulators may have distorted incentives since they are not inclined to take into account the welfare impact of a particular roaming charge on foreign users and foreign operators.

[19] Discounts do exist in practice (e.g. based on volume or destination) but their actual extent is unclear. This does not change the force of the argument: if discounts necessarily apply to everybody, then their effect is diluted.

4.3 Optimal approach

Assume that total welfare is given by the sum of welfare (i.e. consumer surplus and firms' profits) in the two countries affected by an international roaming agreement.

The 'first best' situation involves retail prices set at their incremental cost. As long as retail roaming charges do not include any mark up due to exercise of market power, the implication is that the socially optimal wholesale charges are equal to their underlying incremental cost.

If networks have to recover fixed costs, wholesale charges may include a mark up above costs, but the actual level of this mark up is quite intricate to calculate since it depends on how other prices are set. If mobile subscribers have inelastic demand for subscription (i.e. the market is mature), then there is no reason to distort calling patterns as all fixed costs can eventually be recovered via higher subscription charges rather than via calling charges.

On the other hand, if the market is not mature, then a wholesale mark up would reflect a complicated set of elasticities in both countries. If visiting customers are not very responsive to international roaming prices, then the mark up could be quite consistent since it makes a lot of sense to charge them on this particular service as long as this permits the price of other price-sensitive services to be kept down. The actual level depends on how the roaming retail price affects the subscription decision of foreign users.

This positive mark up can actually be offset by the RPP feature associated with international roaming charges. If, since they are charged for receiving calls, visiting customers tend to keep their phones switched off, then a negative mark up is needed to induce the customers to keep them switched on.[20]

The welfare consequences are quite involved since costs and benefits from two countries should be taken into account. In practice, there is a tendency to neglect the impact on foreign users. If this is the case (but it would not be optimal

[20] RPP is sometimes used as an argument to say that customers are actually quite sensitive to international roaming charges. If this is true, then the force of the previous argument (a wholesale mark up should be imposed to recover fixed costs) is much reduced.

for overall welfare) and a certain NRA was able to set the wholesale charges, it would set higher mark ups the lower the weight attached to foreign users.

4.4 Remedies and their consequences

This section contains a discussion of potential remedies applied to the wholesale national market for international roaming.

Remedy 1: Abolish the no discriminatory clause

No discrimination has received a very restrictive interpretation so far. No discrimination can be appropriate in some situations when single dominance is found in a market. However, it is unlikely that individually dominant operators are found in the wholesale market (although there may be exceptions).

The abolition of this clause seems a 'good' remedy, in the sense that it is not intrusive, and would lead to a market based solution. Visiting operators would have increased incentives to seek better deals as these would allow them to offer cheaper tariffs to their own customers (relative to their rivals). The end result is that wholesale charges could go down (as firms are actively searching for cheaper roaming charges) and these discounts would be passed on to end customers.

On the other hand, this particular remedy would not work if:

a) Technology does not work to direct traffic (as actual demand would still be disconnected to prices and allocated randomly). However this remedy still seems a natural first step: technology would follow once proper economic incentives are in place;

b) Customers are not price sensitive (as price cuts would not obtain any particular impact on users). This is an empirical matter and can be addressed with other remedies, such as promoting better customer information;

c) There is still collusion among operators. There are two aspects that are relevant in this context: a) operators themselves may not jointly want to remove the clause; b) once the clause is removed, whether joint dominance

is affected. With respect to the last aspect, notice that if this clause is removed, it would be more difficult to enforce tacit collusion: a deviant firm would have more gains from one-shot deviations as it could attract a good share of the market;

d) National regulators have conflicting objectives. A country that is a net receiver of international roaming traffic may not want to abolish the no discrimination clause if this leads to reduced charges that could be used to increase national profits and/or national consumer surplus.

Remedy 2: Technology

This remedy would try to induce firms to adopt particular technological solutions in order to increase their ability to direct roaming traffic. This remedy could work since it would re-establish a link between prices and actual demand. However, this remedy is more intrusive than the previous one (as the regulator imposes technological choices) and would probably be redundant if the previous remedy is applied: firms would find themselves applications since they would now have the right incentives to do so.

Remedy 3: Exclusivity

This remedies allows an operator to seek for exclusivity deals, so that roaming customers would be able to connect to a single visited network. This remedy could work as it would induce competition among visited network to attract such traffic (the visited operator that wins the deal gets the roaming traffic with probability one), hence it would cause prices to decrease.

The downside of this remedy is two-fold. On the one hand it may be an inefficient solution if operators differ in their quality of coverage, as a customer may be unable to make or receive calls in some parts of the visited country (this problem is not particularly serious where coverage of visited networks are more or less symmetric – as it is the case in most European countries). On the other hand, exclusivity may lead to some forms of anti-competitive behaviour. For instance a network that is present in many countries and signs most of these deals may get

substantial market power in the segment of corporate users who travel a lot. However, this is a problem that depends on many details (such as the length of the exclusivity deal) and may be tackled by ordinary competition policy.

Remedy 4: Reciprocity

This remedy might help to the extent that there are double mark up problems (and countries are similar). However, since charges are already set in bilateral agreements it may not change much.

This remedy may also help when two separate NRAs are in charge of setting wholesale roaming charges. As discussed above, when a certain NRA sets the wholesale charges, it would naturally tend to neglect the impact on foreign users and companies. This tendency is similar to a standard double mark up problem (rather than neglecting the impact on rival's profits, it is the impact on foreign welfare that is neglected). This latter problem is off set by requiring reciprocal charges in the two countries. However, if a country is a net receiver of foreign users, it is likely that it will prefer high roaming charges, since national welfare would increase (depending on the intensity of national competition, roaming revenues could stay with the operators/shareholders or be passed on to national customers).

Remedy 5: Transparency

This remedy would try to increase transparency of end user tariffs, hence making customers more price responsive. There are no particular downsides, although it is not likely that this remedy alone could fix the problem.

Remedy 6: Imputation tests

Imputation tests such as price squeeze tests would be easy to put in place, linking the prices of national and international calls. However there is no strong economic justification for this type of remedy as demand functions are likely to be very different.

Remedy 7: Charge control

Under this remedy the NRA directly sets wholesale charges. This remedy would obviously work if there is appropriate cost and demand information. However it suffers from some drawbacks. Firstly, it is quite an intrusive remedy in this particular market. Secondly, even if all the information was available, it is not clear why a regulator would want to regulate a wholesale charge that mostly impact on the welfare of foreign users (this is where a some sort of reciprocity between NRAs would be needed). However, assuming there is international co-ordination, this remedy may be necessary if tacit collusion is found in the wholesale market so that less intrusive remedies would not be likely to be very effective.

5. References

Armstrong, M., 2002, "Call termination on Mobile Networks," *mimeo*, Nuffield College and Oftel.

Armstrong, M., 2002, "The Theory of Access Pricing and Interconnection," in: M. Cave, S. Majumdar and I. Vogelsang, eds., *Handbook of Telecommunications Economics*, North Holland, Amsterdam.

Competition Commission, 2003, *Vodafone, O_2, Orange, T-Mobile*, London.

DeGraba, P., 2000, "Bill and Keep at the Central Office as the Efficient Interconnection Regime," OPP Working Paper #33, Federal Communications Commission, Washington (DC).

DeGraba, P., 2003, "Efficient Inter-carrier Compensation for Competing Networks - When Customers Share the Value of a Call," *Journal of Economics & Management Strategy*, 12(2).

European Commission, 2002, "On the Initial Findings of the Sector Inquiry into Mobile Roaming Charges," Working Document, DG Competition, Brussels.

Gruber, H. and T. Valletti, 2003, "Mobile Telecommunications and Regulatory Frameworks," in: G. Madden, ed., *The International Handbook of Telecommunications Economics*, Edward Elgar, Chelthenham, 2003.

Hausman, J., 2002, "Mobile Telephone," in: M. Cave, S. Majumdar and I. Vogelsang, eds., *Handbook of Telecommunications Economics*, North Holland, Amsterdam.

Laffont, J.J. and J. Tirole, 2000, *Competition in Telecommunications*, MIT Press, Cambridge (MA).

Mason, R. and T. Valletti, 2001, "Competition in communication networks: Pricing and regulation," *Oxford Review of Economic Policy*, vol. 17, n. 3, 389-415.

Salsas, R. and C. Koboldt, 2002, "Roaming free? Roaming network selection and inter-operator tariffs," DotEcon DP 02/01.

Valletti, T., 2003, "Is mobile telephony a natural oligopoly?", *Review of Industrial Organization*, vol. 22, n. 1, 47-65.

Wright, J., 2002, "Access Pricing under Competition: an Application to Cellular Networks," *Journal of Industrial Economics*, vol. 50, n. 3, 317-335.