

Regulating technically dynamic industries: the case of 3G call termination

David Newbery
Department of Applied Economics, Cambridge
22 October 2004

Introduction

Regulation of network industries evolved from contractual relationships between investors and municipalities. The US experience is instructive, as it managed to evolve a system for the public regulation of privately owned utilities that survived and prospered for well over a century. Most other countries took such utilities into public ownership, avoiding the discipline of financing investment that requires setting prices high enough to earn a return on capital.

In the US, most cities initially offered a contractual franchise to the new utilities, starting with gas in New York in the 1820s. These contracts typically provided for access to public rights of way and a franchise monopoly in return for restraints on prices and concessional terms for supplying the municipality. The contracts were typically for 20-30 years, long enough to repay the large capital investments. Inevitably, they had to be renegotiated as circumstances changed. Soon contracts made explicit provision for renegotiation, subject to arbitration, or reference to an independent committee, which might also monitor service quality. The power of these regulatory committees grew, and in due course evolved into state public utility commissions. Where competitive entry threatened monopoly profits, as with electricity and telephones, utilities sought franchise protection in exchange for forms of regulation that allowed utilities to pursue their technological objectives. According to Williamson (1985: 347), rate-of-return regulation was a development of an incomplete long-term contractual relationship that offered utilities a fair rate of return in exchange for the ability to renegotiate the terms of the contract without excessively costly haggling.

Rate-of-return regulation evolved through a series of land-mark court cases in the US to provide procedural fairness in the allocation of rents accruing to franchise monopoly investor-owned utilities. It worked quite well for industries where there were no obvious competitive alternatives, no disruptive technical changes, and particularly where gradual improvements allowed the nominal cost of service to fall, removing the need for frequent rate reviews. Thus in the period 1950-73, as the electricity industry moved to larger, more efficient generating sets, and as fuel prices remained static in nominal terms, electricity utilities did not need to apply for rate reviews to raise prices. They thus faced the same incentives to cut cost as if their prices were capped, for they could rely on retaining the benefits of cost reductions for a considerable time until a rate review became necessary.

Two events disturbed this satisfactory equilibrium. The first was the oil price shock of 1973, which set in train strong inflationary pressures, fuel price increases, and a fall in demand growth that undermined the regulatory compact in the US electricity sector. The second was

rapid technical progress in long-distance telephony (particularly in switching and the feasibility of micro-wave transmission) that exposed the growing cross-subsidies from long-distance to local telephony. Entry and the threat of by-pass caused AT&T to attempt to rebalance its charge structure, thereby attracting Department of Justice concerns of predatory pricing. The legal battles that unrolled over the next decade were only resolved when AT&T proposed unbundling, accepting that long-distance telephony was a potentially competitive service, no longer a natural monopoly requiring a regulated franchise.

When Britain considered privatising its own network utilities (starting with British Telecommunications – BT - in 1984) it had to consider how they should be regulated. Initially it was proposed to subject BT to rate-of-return regulation, but Alan Walters, the Prime Minister's advisor, strongly criticized this. He argued that rate-of-return regulation effectively imposed a 100% tax on all profits in excess of the allowed rate, with terrible incentives for efficiency and innovation. Stephen Littlechild was commissioned to advise on regulation, and recommended a local tariff reduction scheme, rapidly renamed RPI - X. His idea, set out in Littlechild (1983), was to confine regulation to the non-competitive core (local service), and leave competition to restrain other prices, thus providing incentives for efficiency and avoiding the complexity of the US system of regulation.

The other great merit of a price cap was that it mimicked the effect of a competitive market, and would serve well as a temporary restraint until competition developed. He therefore did not give much consideration to how such price controls might be reset, nor how one might distinguish between segments of enduring market power that would continue to need regulatory oversight, and those segments that could confidently be left to the market.

The theory and practice of regulation has come a long way since 1984, the year in which BT was privatised and AT&T unbundled. Regulating natural monopoly networks in electricity, gas, water and rail is now accepted as an enduring requirement for satisfactory private ownership that can sustain both political support and deliver investment to maintain quality of service. These networks are alike in being capital-intensive, highly durable, and experiencing little or modest technical progress. Price-cap regulation has evolved principles for establishing prices that provide investors with assurance that their investments, if undertaken efficiently, will deliver an acceptable return. The main thrust of efficiency gains to date (the X factor in RPI - X) has been to move less efficient firms towards the efficiency frontier, and to realise economies of scale with growing demand. Predicting future technical progress is difficult, and regulators are instead increasingly attracted to yardstick regulation, in which each company is rewarded to the extent that it can outperform its peers. The US in turn is increasingly attracted to performance-based or incentive regulation, in contrast to cost-of-service regulation such as the classic rate-of-return system. Adapting to the market-oriented philosophy of price-cap regulation is resisted, but increasingly its merits are being recognised (Taschdjian, 2000).

As more networks are unbundled and privatised in an increasing range of countries, so regulators are accumulating evidence and insights into the challenges and limitations of regulation. The evidence that regulation is costly is now widely accepted, and was the driving force behind the deregulation movement in the US. The cost does not just include the failure to adopt existing best practice where it would reduce costs, but also the various inhibitions to

innovation in any cost-based regulatory system. Innovation typically requires up-front investment for a service that does not yet exist, and for which there is no obvious source of finance where all prices are regulated. The company would have to argue for a revenue stream to finance this up-front cost, and the regulator would then be faced with a dilemma. Clearly the bright-eyed engineers would like to have finance for their pet ideas, so the regulator must judge whether the investment is in the consumer interest before effectively charging consumers now for the future delivery of an untried new service. But regulators lack the commercial insights and judgement to make that decision, and are likely to hesitate before authorising more than modest sums.

Hausman (1997) argues that one of the most costly consequences of traditional US-style regulation is the delay imposed by regulators concerned that the new services might be cross-subsidised from the captive customer base. Hausman gives two examples - voice messaging and cellular phones. Cellular (i.e. mobile) phones were apparently sufficiently advanced to have been introduced in the early 1970s, but regulation delayed introduction until 1983. Hausman estimates that the loss of consumer surplus from delaying the introduction of cellular phones might have been as high as \$25-50 billion per year, and the total cost of delay possibly \$100 billion. This is large compared with the 1995 US global telecoms revenues of \$180 billion/year.

A market-oriented approach to regulation is likely to take a different view, in that new services would normally be unregulated, as they would not appear in the basket of services that forms the price cap. Investors would risk their shareholders' money when investing in new services, and would expect to be free to set prices on an unregulated market. If the service is successful, the shareholders (and consumers) win, and if not shareholders lose, but consumers have not been forced to cross-subsidise the investment out of regulated services. It is difficult to see how else to treat new services in technically dynamic industries.

Regulating mobile network operators

Mobile telephony began to have a significant market impact in Europe with the introduction in the early 1990s of GSM (Global System for Mobile Communications, in this version often called Second Generation, or 2G mobile telephony).¹ In 1990, penetration for the EU-15 was less than 1 per 100 inhabitants, while for the UK it was just 2. By 1995 penetration was still below 5% (at 4.8 per 100 inhabitants) for the EU-15, with the UK again nearly twice that, and the Nordic countries already over 20%. The market subsequently took off at about this time, when the EU-15 began to rapidly catch up with the UK (reaching 14% penetration, compared to the UK's 15%, by 1997). By 2000, the UK and the rest of the EU were rapidly catching up with the Nordic leaders. The UK had a penetration of 73%, the EU 64%, compared with Sweden and Finland on 72%, almost double the US with a 2000 penetration of 40%.² Growth continues, with UK's 2002 penetration at 84%, Italy's at 93% and Luxembourg with 101%, with many markets thus nearing saturation.

¹ 2G uses spectrum within the around 900 or 1800 MHz.

² Data on penetration from Eurostat (2002).

During this period of rapid growth, the industry was not subject to any price regulation. It fit the description of a technically dynamic and risky new industry with a considerable prior history of very modest performance (particularly the early years under analogue coverage from 1985). Most companies were struggling to make profits given the very substantial costs of rolling out the network – perhaps as much as 40-50 billion euros for the EU-15 by 1998.³ Mobile networks provided facilities-based competition to fixed line telephony and on the accepted British regulatory logic, would reduce the need for costly regulation. Competition between MNOs became intense once sufficient licenses had been issued, and the MNOs competed fiercely by subsidising hand-sets and cutting calling charges to build up a customer base and realise the economies of scale in the network.

In such a market environment, innovations in design, services (such as SMS), and pricing (particularly pay-as-you-go), were critical for creating rapid growth, and were rapidly imitated. The need to finance the hand-set subsidies (and other start-up costs) induced MNOs to set higher termination charges than might have been warranted on a simplistic cost-allocation approach, but the resulting subsidies to customer acquisition internalised the network externalities that made large customer bases attractive to new subscribers and help stimulate market growth.

In 1998 the British Director General of Telecommunications made a reference to the Competition Commission on the charges made by Vodafone and Cellnet for terminating calls from fixed-line networks. The Competition Commission subsequently found that

“We conclude that there is currently insufficient competitive constraint on termination charges.

“We note, however, that mobile telecommunications is still a relatively immature service in the UK. The industry and the technology are evolving rapidly and new competitive pressures may emerge. The client base is growing in size and experience and it is possible that charges for incoming calls will assume greater competitive significance in future. In these circumstances we do not believe we have a sufficient basis to expect that competitive constraints on termination charges will remain inadequate beyond the next three or four years.” (Competition Commission, 1998, 1.5-6).

The decision is noteworthy for several reasons. First, it was confined to two of the four MNOs (who at that time had over three-quarters of the total market) and by that measure

³ Eurostat (2002) estimated the investment cost of total ICT equipment for the EU-15 at 133 billion euros in 1998, and 173 billion euros in 2002, or 460 euros per head of population. This includes computing, fixed line as well as mobile, and so it overstates the cost of mobile network investment, but it is plausible that perhaps one-quarter or 44 billion euros per year can be attributed to mobile network investment. ITU (2003) gives mobile revenues for the EU-15 in 2002 as \$100 billion, 35% of total telecoms revenue for that year. Total telecoms investment for the EU-15 (less Greece) was \$49 billion, and a considerable fraction of that, perhaps three-quarters, must be mobile investment, as fixed line telephony is growing very slowly.

had significant market power. It did not apply to the two entrants. Second, regulation was confined to the market in which they had a demonstrable monopoly (call termination), and even here, it only applied a price cap on call terminations from fixed line, not on terminating mobile-to-mobile calls. Third, it was time-limited, with any continuing regulation requiring a new market assessment. Finally, the proposal to regulate was not made until penetration had exceeded 9 million subscribers (Competition Commission, 1998, 4.3).

Consequences of regulating mobile network operators

Mobile telephony had been originally licensed with no indication that it would be subject to economic regulation, as opposed to normal competition law.⁴ Clearly, Oftel's 1998 proposal, upheld by the Competition Commission, changed the rules of the game. Effective regulation requires the trust of all parties (regulated and consumers), and that it turn requires that it is based on sound reasoning, is justified and proportionate. Here the important clarification was to further articulate the concepts of market definition and of significant market power (SMP). In classic cost-of-service regulation, the emphasis was on costing the individual network elements, and this led to market definitions based on technology or elements (local loop, trunk, fixed, mobile, etc). The more market-oriented approach to regulation adopted in Britain and the EU instead defined the relevant markets by service (call origination, call termination).⁵ This was subsequently clarified in EC Communications Directives, which came into effect on 25 July 2003. The Framework Directive defined Significant Market Power in Article 14 to be equivalent to the competition law concept of dominance.

On this reasoning, call termination was a market in which MNOs had SMP, as the MNO to which the customer subscribed was the only supplier of call termination services to that customer, and therefore had a dominant position (100% of the market of terminating calls to that customer). While the principle of imposing *ex ante* regulation such as price caps to address cases of SMP was reasonably clear in theory, the practical issue of how to set the price caps was anything but clear in practice. There were disputes about the details of cost modelling, disputes about the empirical size of the fixed and common costs that needed to be allocated to the various markets or services, and even more fundamental disputes about the determinants of the mark-ups to apply to the Long Run Incremental Costs (LRIC) of call termination. Article 13 (2) of the EU Access Directive (2002/19/EC) requires "National regulatory authorities shall ensure that any cost recovery mechanism or pricing methodology that is mandated serves to promote efficiency and sustainable competition and maximise consumer benefits. ..." The problem lay in how to interpret the promotion of efficiency and the maximisation of consumer benefits. The MNOs, in their submissions to the Competition Commission, argued that both requirements implied Ramsey pricing, while the Competition Commission claimed that consumer benefits would only be maximised if all unregulated

⁴ The Telecommunications Act 1984 does empower the Director General to modify licence conditions, and to make references to the Competition Commission under Section 13 on matters that "operate, or may be expected to operate, against the public interest."

⁵ See Taschdjian (2000) for a fuller comparison of the differences between the two approaches.

markets (i.e. for access and origination) were perfectly competitive, which it disputed. In any case it felt that the empirical evidence required for setting the Ramsey mark-ups was lacking.

Third generation mobile

The next or third generation (3G) of mobile phones based on UMTS (Universal Mobile Telecommunications System) uses higher frequencies than 2G. The new technology allows a range of additional services, and greatly expands the band-width available (which will be needed for the enhanced services and also to meet the high level of normal voice demand more efficiently). Now that 3G spectrum has been allocated or auctioned, and networks are being constructed, National Regulatory Authorities (NRAs) are faced with the question of how to treat 3G voice termination when in most countries 2G voice termination is subject to a price control. The market analysis required under the EU Communications Directives will surely find that 3G MNOs' have SMP in the market for call termination on 3G networks, and for exactly the same reason that MNOs are currently deemed to have SMP for call termination on 2G networks. Indeed, some NRAs (notably in Ireland, Greece and Sweden), are already proposing to extend regulation that currently applies to 2G voice termination to 3G voice.

Neither the issue of market definition nor that of SMP is under dispute here; the question addressed here is whether their presence necessarily requires NRAs to impose price-cap regulation on 3G call termination. At this point we need to return to the undesirability (and difficulty) in regulating technically dynamic industries, and the provision of new services. 3G call termination is in one sense a new service (as it enables new forms of communication to be provided and delivered), while at the same time being a substitute for an already regulated service (2G voice call termination). The case for regulatory forbearance therefore rests on a balance of considerations, of which the most important are the following.

- First, there are major risks of introducing regulation in the early stages of the investment cycle for the service because of the impact on innovation incentives. 3G services are new. Investments are still being made in building and rolling out networks and services are yet to be launched in most countries. The take up of these services are uncertain in the early years of their deployment. The demand for these services and the associated investment required is uncertain. The European Commission has itself recognised this in its discussion of emerging markets⁶, and it is generally recognised that firms can and should

⁶ The distinguishing feature of an emerging market is that the market is immature. Even if a firm makes non-trivial investments to be able to provide a new service there is no guarantee that, in an innovative and fast moving sector, a cheaper alternative mechanism for delivering the service will not be found. It is also difficult to assess the dynamic of competition behind any entry barrier, as many potential entrants will not make firm plans to enter a new service area until the market is seen to be a commercial proposition. Many new initiatives on the marketplace fail but successful ones create incentives for other firms to enter the market. In discussing the second criteria, in the Explanatory Memorandum to the Recommendation, it is stated that "entry barriers may also become less relevant

be allowed to capture first mover advantages in markets without fear of immediate regulation – if regulators step in too early and expropriate all the gains from innovation then firms are unlikely to assume the risk in the first place. There is therefore much debate as to when markets become sufficiently mature for first mover advantages to turn into monopoly profits needing regulation.

- Second, there are large risks of regulatory error when applying regulation to new and innovative technologies. Regulators are poorly equipped to make commercial judgements and to undertake social cost-benefit studies to see whether introducing a new technology is justified (or how much “excess” mark-up over short-run costs such technologies should be allowed). We have already noted that US regulators inhibited the introduction of voice messaging and cellular phones (Hausman, 1997). Tashdjian (2000, 5.3) observes that German and Dutch requirements to sell special-access facilities at prices based on LRIC led to “a dramatic drying up of local infrastructure investment”. Similar disputes have surrounded the pricing of unbundled local xDSLs and may have impeded the competitive supply of broadband in Britain and elsewhere.
- Third, there are a number of more practical problems that regulators will face if they decide to formally regulate 3G services. One option is to recalculate cost models with both 2G and 3G costs. Some NRAs have argued that 3G call termination rates should be set equal to the 2G price cap, and so avoiding the need for a more complex cost modelling exercise to account for 3G costs. Other regulators have decided to leave 3G call termination unregulated at this stage and adopt a wait and see approach. Each of these approaches has potential attractions and drawbacks, which need to be understood in order to make a sensible regulatory decision. The choice could have far-reaching consequences, for it is most unlikely that 3G represents the final stage of development of mobile telephony, particularly in a world of growing convergence of different forms of ICT. Already wi-fi and more intelligent uses of spectrum hold out the prospects of networks capable of combining the benefits of mobility with the power of more conventional computing.

Poor regulatory decisions could inhibit investment in uncertain and costly developments, raise the cost of finance to the industry, and delay the experiments needed to validate speculative concepts. Good regulation should be predictable, credible, based on widely accepted principles, and the minimum needed to prevent costly inefficiencies or

with regard to innovation-driven markets characterised by ongoing technological progress. In such markets, competitive constraints often come from innovative threats from potential competitors that are not currently in the market. In such innovation-driven markets, dynamic or longer term competition can take place among firms that are not necessarily competitors in an existing “static” market.” It is only with the elapse of a sufficient amount of time that these questions can be answered. (ERG “Common position in the approach to remedies in the new regulatory framework”, p.21-22)

customer exploitation. In the words of the EC Communications Directive, any regulation “shall be based on objective, transparent and non-discriminatory criteria and shall be consistent with the principles of proportionality and of least market distortion”, i.e. they must be justified and proportionate and market friendly. With these criteria in mind we can now examine the various alternatives suggested.

1. Regulating 3G termination charges at the 2G level

Some NRAs argue that 3G mobile networks deliver voice much as 2G networks, and although they may also permit the supply of other services, these are irrelevant for the purposes of regulating voice termination. MNOs would, they argue, not have invested in 3G networks unless they could deliver voice services at lower cost than 2G (perhaps after receiving credits from the profits of providing these other services). Given the difficulty of actually assessing 3G costs (specifically, the cost of spectrum and the correct interest and depreciation rates to apply), the simplest solution is to regulate 3G termination at the same level as 2G charges. Thus the Swedish National Post and Telecom Agency (PTS) in Appendix 2 of its draft entitled “LRIC Pricing Method for Termination of Voice Calls in Mobile Networks” dated 10 May 2004 states that:

The PTS draft obligation decision indicates that no distinctions are made between voice calls in GSM networks and termination of voice calls in UMTS networks. It was already made clear in the Mobile Calculation Project's Guidelines for Model Work (Conceptual Design) that the model would be designed in order to reflect certain migration of voice traffic from the GSM network to the UMTS network but that costs for the UMTS network would not be modelled explicitly due to a great deal of uncertainty and lack of availability of experience or necessary data in order to be able to design such a cost model. Therefore, the model calculates the cost of terminating voice calls in the GSM network.

The Conceptual Design document (page 39) states that UMTS is anticipated to be a system involving greater efficiency in terms of the cost per traffic unit being lower in UMTS than in GSM, even if this cannot be proven beforehand. In addition, it is stated that GSM is indicated as "Modern Equivalent Technology" for this service, i.e. that GSM today is the most modern and accepted technology in which to terminate voice calls in mobile networks. ...

It would not be deemed as efficient to terminate voice calls at a higher long term cost. Therefore, it can be assumed that the cost which can be compensated through the termination compensation in the UMTS network must be lower than, or at the most equal to, the cost of the corresponding service offered with the established modern technology (Modern Equivalent Technology) for termination of voice calls in mobile networks: GSM.

PTS believes it is reasonable not to consider migration to the UMTS network in the model since no distinction is made in the obligation decision between 2G and 3G termination. In light of this, it is also reasonable that the cost for termination of voice calls in 3G networks is not permitted to be higher than in GSM.

The arguments here need teasing out. The first argument is that “no distinctions are made between voice calls in GSM networks and termination of voice calls in UMTS networks”, suggesting that termination charges should be technology neutral. This could lead

to the second proposal that charges should be based on a mixture of the costs of the various technologies offered. In Britain, the different costs of 900 and 1800 MHz call termination are recognised and MNOs with different spectrum allowed different termination charges, so other interpretations are certainly possible. The second argument is that “that costs for the UMTS network would not be modelled explicitly due to a great deal of uncertainty and lack of availability of experience or necessary data”. This suggests that PTS was contemplating setting termination charges at some suitable average of the costs of 2G and 3G but was handicapped by lack of reliable 3G data.

The third argument is that “UMTS is anticipated to be a system involving greater efficiency in terms of the cost per traffic unit being lower in UMTS than in GSM, even if this cannot be proven beforehand.” In other words the MNOs reveal by their willingness to pay for spectrum that the new technology is profitable and presumably therefore has lower costs than the existing 2G. This argument is not completely convincing as there is a shortage of spectrum for continued expansion of 2G, and the spectrum scarcity price is not included in the present 2G LRIC estimates. However, it seems an acceptable hypothesis, although there continues to be uncertainty about its validity.

The final argument is that “it is stated that GSM is indicated as ‘Modern Equivalent Technology’ for this service, i.e. that GSM today is the most modern and accepted technology in which to terminate voice calls in mobile networks.” That argument is the strongest for basing 3G termination charges on current 2G costs.

Problems with setting 3G termination at 2G levels

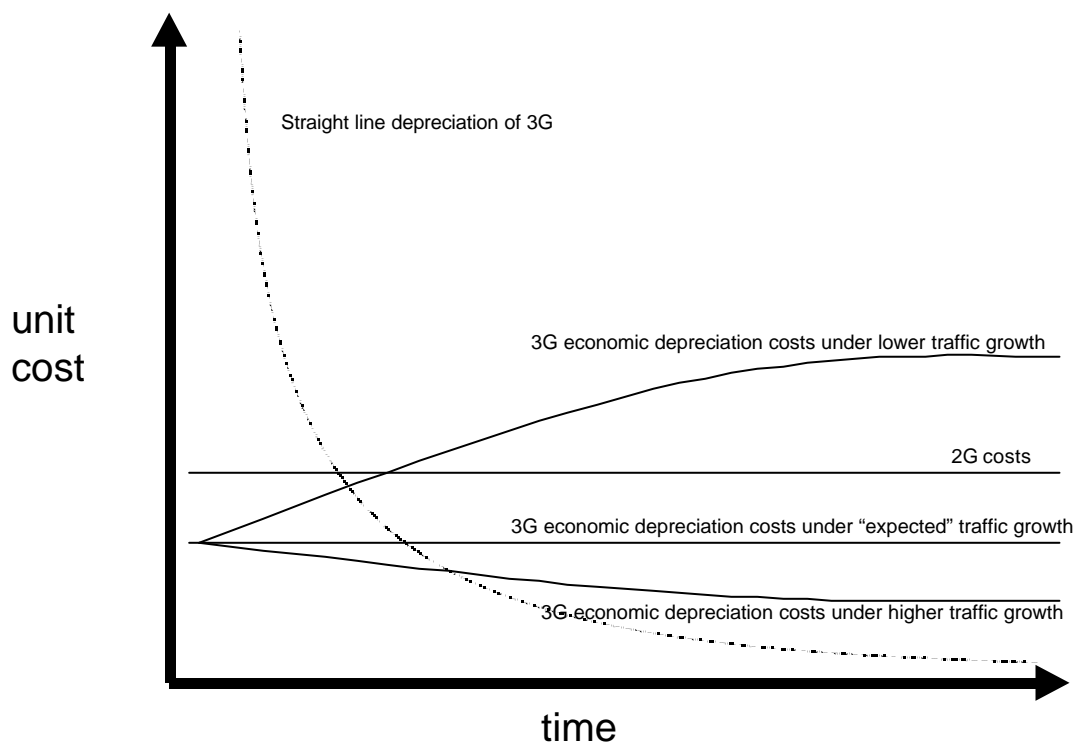
The main problem with this line of reasoning is that it is based on the concept of a modern equivalent technology. That being the case, MNOs will anticipate that at some future date once 3G has reached sufficient penetration and maturity it will become the new modern equivalent technology. At that point NRAs will set 3G termination rates on the basis of the current (lower) 3G costs, to pass on the benefits of technical progress to consumers. While it may be true that a fully used 3G network would have lower unit costs than a fully used 2G network (*inter alia*, because it uses spectrum more efficiently), there are large fixed costs in establishing the network that imply that average costs will exceed average 2G costs until some critical customer base is reached. During the period before this critical penetration is achieved, revenues based on 2G costs will be insufficient to pay for the 3G network. If the charge is set at the 3G average cost level at the first moment that this critical penetration is achieved, then the MNOs will incur losses up until that moment which they would not recover thereafter.

NRAs may argue that the problem as identified is an issue of setting the correct economic depreciation profile. Under straight line depreciation unit costs of 3G traffic will be very high in the early years. However, an economic depreciation profile calculated on expected traffic or revenue growth could ensure an equalized unit cost throughout the period (or perhaps tilted to reflect declining market prices for a unit of output over the asset life). This equalized cost, it could be argued, should be lower than the present 2G costs, if the

former is more efficient. On the face of it, this would seem to justify taking 2G costs as a price ceiling, but switching to a lower 3G cost base once such a cost base can be established.

However, there are at least two problems with this. The first is simply one of establishing the correct “regulatory compact”. The economic depreciation profile (that would tilt depreciation according to expected traffic or revenue growth) would need to be agreed by the company with the NRA as part of a “regulatory compact”. There would need to be some means of assuring the regulated company that it would be allowed to pursue the tilted depreciation profile throughout the asset life, recognizing a much higher depreciation charge in later years. There are some examples in capital-intensive industries where the time-profile of the price cap is engineered to ensure that various financial ratios remain within acceptable limits. These can also be readjusted at periodic reviews to ensure that the operators receive an adequate return on and of their efficiently incurred investment, but these examples all come from technically fairly static industries such as water and gas transmission, where the investment is a large part of the cost of the business.

The second problem is more fundamental, and is related to difficulties if actual demand falls short or exceeds the expectation on which the economic depreciation profile is calculated. This could result in actual unit costs being significantly higher or lower than expected in the later years of the asset life. This is illustrated in the graph below.



This would introduce a significant asymmetric risk to the company. To see this consider the two cases of where actual demand: (1) exceeded expectations; (2) fell short of expectations. If the first case occurred, unit costs would be lower, and the NRA would wish to capture these benefits for consumers in any price cap revision. In the second case the

regulator might wish to enforce the 2G price cap (as evidently the most efficient technology). However, since the agreed 3G depreciation profile required the bulk of the costs to be recovered in the later period, a large proportion of the 3G investment costs would then go unrecovered.

It is clear, therefore, that the risks perceived by the company are asymmetric. If demand exceeds expectations, the benefits will be handed back to consumers. If demand falls short of expectations the company will need to bear the shortfall. This would result in a significant disincentive to invest in a technology that could be expected to yield benefits to consumers.

Effectively, the regulator is requiring the company to grant a real option to consumers such that they benefit from lower future prices made possible by the 3G network, should demand be sufficient. In principle the value of the option (or cost to the company) can be computed from an assumption about the variability in the future demand forecast. It becomes clear, therefore, that there is an additional cost to the company over and above the cost of the existing 2G network, and so the 2G price cap will cease to be correct under this regulatory regime.

In short, this approach to regulating 3G lacks the credibility that it will be proportionate and justified, in that the costs incurred in establishing the network and which lead to higher initial average costs than the existing 2G network risk not being compensated.

2 Regulating 3G termination on the basis of 3G LRIC

This second approach seems to attract the PTS, but is rejected as impractical as the necessary data to populate a cost model is too uncertain or is just not available. Quite apart from anything, to be practical it would probably amount to the detailed cost-of-service regulation that bedevils US regulation (and which Hausman and Taschdjian have criticised so cogently). If costs are very uncertain, then prices would have to be adjusted to reflect the costs that were “prudently incurred”, and the regulator would be faced with the impossible task of trying to assess whether (or to what extent) the innovation was prudent – i.e. socially beneficial. This may work with established technologies such as traditional electricity generation (but failed spectacularly with nuclear power in the 1970s, as Gilbert and Newbery (1994) document). It has poor incentive properties, leading to gold-plating if all costs can be recovered from consumers. Faced with that risk, regulators have to judge what the market will value, a difficult task that they are ill-equipped to undertake. That is precisely the reason for allowing innovators to take the risk but also to reap the reward if successful. To take another example, the patent system is designed to ensure not only that inventors can exploit their innovation but that imitators cannot cheaply replicate their ideas and compete. Not only are the innovations not regulated, but inventors are allowed a period in which they can market their innovation with a protected monopoly.

If regulators were nevertheless committed to regulating 3G on the basis of 3G costs, they would have to make two difficult calculations: of the option value of the investments, and of the rate of depreciation of the assets. If done correctly, these two adjustments to conventional cost accounting would likely considerably increase the relevant costs and hence

raise the price cap, possibly even above the 2G price cap. Real option theory deals with irreversible sunk investments in risky technology, where the future demands or costs are uncertain, but where investors can learn from experience and the passage of time. In such cases, MNOs contemplating an investment decision have the option to “wait and see”, during which they can undertake market research, observe the evolution of demand and technology, and learn more about the economics of the investment, or to make the investment and thus lose the chance of any subsequent learning. The option value of waiting to learn more can be large if the uncertainty that may be resolved is large. Once the investment is made this option value is lost, and so the rule is to invest only when the expected net present value of the investment is larger than the real option value. It is the sum of the investment and the option value that is relevant for determining the regulatory asset base in any cost-based calculation of LRIC, and this amount is not readily observable in the market (unlike financial options, which are often traded).

Even if the correct capital value can be estimated (and there are formula for valuing real options), the LRIC requires an assessment of the return on this capital value and the depreciation of that value. The economic depreciation is hard to estimate, for several reasons, one of which (the market size) has already been mentioned. In the case of technically dynamic industries, the future asset value will depend on the future cost of replacing or substituting the services it supplies, so the fall in value (i.e. the economic depreciation) will depend on the rate of technical progress. For example, computers subject to Moore’s Law and falling in unit cost by 50% every 18 months will have to write off 37% of the opening value every year. The charge for interest and depreciation will be higher the faster is expected technical progress, and clearly can be very high in the early years.

The risk from the investors point of view of this form of regulation is that the regulator may also “wait and see” before determining the allowable costs for the LRIC determination. If there are any asymmetries (if costs turn out to be low then price caps are revised down, but not revised up in the event that costs turn out to be higher), then investors will rationally delay or require higher hurdle rates of return for the whole business (i.e. higher prices for supplying unregulated services), to the detriment of consumers.

It could be argued that compensation for these (asymmetric) risks are already embedded in the cost of capital of the mobile companies, and that cost-based regulation of 2G and 3G termination rates, using a current estimation of the cost of capital, will provide the required incentive to invest. Regulators most commonly estimate the cost of capital from an application of the Capital Asset Pricing Model or CAPM. This measures risk from the company’s “beta”, the degree of covariation of the expected future returns of the company with those of the market, on the assumption of symmetric and normally distributed risk. In practice, the expected future beta is proxied by the observed historical beta over recent years. There are, therefore, two reasons why the CAPM will not pick up the degree of risk associated in 3G investment. First, on a theoretical level, the CAPM’s assumption of symmetric risk is crucially violated, making the CAPM theoretically unsuitable for this element of risk. Second, the normal methods of estimating beta from the historic covariation in stock price returns will not capture the forward looking risk anticipated by investors. If

regulators were really determined to regulate 3G services at cost, they would require a considerably more sophisticated method for dealing with risk than they have been using to date. It is hard to see how they could measure a change in future risk arising from a change in regulatory practice (that had not yet happened when the 3G investment decisions were being taken) when in the past the expectation has been that new innovations will not be regulated.

3 Regulatory forbearance: leaving 3G termination to *ex post* competition remedies

The final approach has been accepted as the only logical solution by the British telecoms regulator, Oftel, in its submission over 2G voice call termination charges, and upheld by the Competition Commission (2003).⁷ To quote from the Ofcom (the successor to Oftel), statement of June 2004:

5.27 The December consultation proposed no *ex ante* regulation of 3G voice call termination services.

5.28 At the time of writing, there is still only one MNO ('3') offering voice call termination over a 3G network. Vodafone and T-Mobile have so far restricted 3G services to data, whilst Orange and O2 have yet to offer 3G services.

5.29 The reported number of subscribers to '3''s services – and thus the total number of subscribers using 3G voice services - in the UK by the end of March 2004 was in the region of between 384,300 and 420,000. This amounts to approximately 0.75% of the total mobile subscribers in the UK.

5.30 At such an early stage of roll-out, the costs of 3G voice call termination are unclear, and robust cost information is difficult to ascertain. Thus, in terms of the charges set for 3G voice call termination, there is currently insufficient evidence to conclude that such charges are excessive.

5.31 Ofcom also considers that any adverse effects to consumers associated with charges for 3G voice call termination are likely to be small, given the very limited size of '3''s mobile subscriber base relative to the wider mobile sector. In Ofcom's view, the lack of evidence of excessive charging, combined with the modest effect any charges have on consumers as a whole,

⁷ The New Zealand Commerce Commission has also come to the preliminary conclusion that regulation should not be placed on 3G termination. See the press release of 18 October 2004 at http://www.comcom.govt.nz/publications/display_mr.cfm?mr_id=1419. In particular, the Commission states "...as the earlier discussion of investment risks makes plain, the scope and size of dynamic efficiency detriments are primarily driven by the risks to the incentives for investment in 3G roll-outs. Yet the 3G networks have not yet been built while the competition problem with which the Commission is primarily concerned is the existence of market power in the provision of termination on the existing mobile networks. the Commission believes that regulation of future 3G networks is likely to increase the risk of delay, or restrict investment in 3G networks and cause a significant dynamic efficiency detriment."

mean that it would be disproportionate to impose *ex ante* obligations on 3G voice call termination at this time. Ofcom does, however, intend to keep this position under review, and will retain the ability to bring forward proposals for regulation if warranted.

5.32 Ofcom therefore remains of the view that no specific *ex ante* regulation of 3G voice call termination services is at present required.

Several issues are addressed in this brief extract. The central conclusion is that Ofcom maintains its position that there is no justification or need for *ex ante* regulation of 3G voice call termination. This is because there is no evidence that such termination charges are excessive, and because in any case the number of affected customers is so small. Ofcom recognises that it would be very difficult to establish a robust estimate of costs in such an immature and dynamic market. Bearing in mind that as normal competition remedies for abuse of market power remain, it is reasonable to expect that unregulated 3G termination charges will be set at a defensible level (specifically, related to costs that the Competition Authorities could estimate robustly). If there were evidence of an abuse of market power, then the normal competition policy remedies of investigation, possible fines, and price control would then apply. In short, while 2G call termination would be subject to classic *ex ante* price regulation, 3G call termination would continue to be subject only to classic *ex post* competition policy remedies.

This seems a very logical treatment of an innovation in a regulated but technically dynamic industry. It addresses the problem of regulatory opportunism, the “catch-22” in which investments that successfully lead to cost reductions are only compensated with their actual costs, with the benefits passed on to consumers, while unsuccessful investments that fail are paid for entirely by the companies. Although consumers appear only to win in this arrangement, companies will be overcautious in investing in risky potentially cost-reducing innovations, thus denying the consumers some share of the potential gains. In effect, the regulator, in this case Ofcom, is saying that rather than regulate any new innovation that substitutes for an existing regulated service, and thus raise the risk that the investor will only recover costs if successful and otherwise stands to lose, we will exempt this innovation from regulation but subject it to the same competition policy rules as other market activities. Once the uncertainty has been resolved, and the new technology becomes the *de facto* industry standard, it can then be considered using the same criteria of SMP as other established technologies.

There are, however, potential difficulties that should be recognised and which might usefully be addressed even with this apparently liberal approach. Again, the problem arises in determining what would constitute an abusive price for the 3G termination element, even in an *ex post* competition inquiry. The problem, as before, lies in valuing the hidden real option value that is lost once the irreversible network investment (and acquisition of spectrum, if this continues to be non-tradable) is made. Given that the investment is risky, this value can be high and justifies a high hurdle rate of return to the investment. Once the investment is made, if unsuccessful, there is no need for a competition investigation, and the MNO loses its

money. If very successful and the market takes off, rapidly dominating the older 2G technology, then regulatory scrutiny will likely be directed at the “excessive” profits that may be earned. The expected return *ex ante* may have been of a 50% probability of a loss of £5 billion on an investment of £10 billion (followed by exit), and a 50% probability of a continuing annual profit of £3 billion. The *ex ante* expected rate of return would be about 10%, but in the event of success it would appear to be 30%, and perhaps deemed abusive, requiring price regulation.

There are standard ways of determining the hurdle rate of return for investments which uncover information (in this case, whether or not the new product would succeed in the market) but they require some judgement about the nature of the uncertainty (the probability distribution of outcomes, and an estimate of the value of each of the possible). These require judgements that the investor is forced to make, but which are difficult and always open to dispute. They are not judgements that can usefully be made with the benefit of hindsight, and it is most implausible that any NRA would be able to make them in advance of the decision. They urge caution in any premature reference to the competition authorities, and the need to establish the expectation that innovations will be allowed an adequate period of time and the chance to secure a substantial market share before any regulatory investigation is launched (as was the case for the earlier generations of mobile telephony).

Conclusions

- Dynamic markets such as mobile telephony require a very different regulatory approach to that traditionally used in stable network industries, where costs and required returns to attract investors can be reliably estimated on the basis of past performance and incremental updating.
 - For industries (like mobile telephony) in which regulators are faced with major innovations in the delivery of the regulated service the approach taken by the regulators must be proportionate if investment incentives are to be protected and innovation encouraged.
 - Without a measured approach, regulators run the risk of creating an environment of asymmetric regulatory risk or bias, where successful commercial services are regulated at cost-orientated prices and unsuccessful commercial services ignored. This could seriously lower the expected return to uncertain innovations, to the point that they may be discouraged. At the very least this would require that any attempt to regulate prices be based on a suitable risk-adjusted WACC if firms are to be able to earn returns to justify investment in innovative products and services. The difficulty of estimating a suitable WACC (and asset value) means that the prospect of such regulation is still likely to chill innovation.

With the transition to 3G networks, there are a number of ways to approach the issue of voice call termination.

- Regulating 3G termination at 2G levels might appear logical, as new technologies should be profitable at the prices justified by the previous, inferior technology. It lacks credibility, though, as eventually, if successful, the new technology will become the industry standard and prices are then likely to be rebased on 3G costs. This might not matter if in the early years 3G costs were lower than 2G, but that is unlikely to be the case, given large set-up costs and modest rates of initial market penetration.
- Regulating 3G termination on the basis of 3G costs is impractical at present, given the difficulty of estimating these costs accurately and allocating them properly over time, and in any case faces the major difficulty of determining the WACC for uncertain innovations.
- Forbearance in the early years of 3G service deployment appears to be the most sensible way to approach the issue as the industry transitions to new 3G networks. This is because 3G services are in the early stages of commercial deployment and demand and associated investment requirements are still uncertain.
- This approach avoids the need to model the costs of 3G networks (a difficult and time consuming task) and better promotes investment incentives compared to simply choosing to regulate 3G voice termination at the same LRIC rate as 2G voice termination.

References

- Armstrong, M., Cowan, S. and Vickers, J. 1994. *Regulatory Reform - Economic Analysis and British Experience*, Cambridge: MIT Press.
- Commerce Commission (2004) *Schedule 3 Investigation into Regulation of Mobile Termination, Draft Report*, 18 October 2004, New Zealand Commerce Commission
- Competition Commission (1998) *Cellnet and Vodafone: Reports on the references under section 13 of the Telecommunications Act 1984 on the charges made by Cellnet and Vodafone for terminating calls from fixed-line networks*, available from http://www.competition-commission.org.uk/rep_pub/reports/1999/421cellnet.htm#full
- Competition Commission (2003) *Vodafone, O2, Orange and T-Mobile: Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks*, available from http://www.competition-commission.org.uk/rep_pub/reports/2003/475mobilephones.htm
- Eurostat (2000) *Information Society Statistics*, Eurostat, European Commission
- ITU (2003) *World Telecommunication Development Report 2003*, Geneva, International Telecommunications Union
- Gilbert, R.J. and D.M. Newbery (1994) 'The Dynamic Efficiency of Regulatory Constitutions', *Rand Journal of Economics*, Vol. 25 No. 4, 538-54,
- Hausman, J.A. (1997) 'Valuing the effect of regulation on new services in telecommunications' *Brookings Papers: Microeconomics*, 1-38
- Littlechild, S.C. (1983) *Regulation of British Telecommunications' Profitability*, London: HMSO
- Taschdjian, Martin (2000) *From open networks to open markets: how public policy affects infrastructure investment decisions*, Harvard University Center for Information Policy Research, Program on Information Resources Policy, Nov.
- Williamson, O.E. (1985) *The economic institutions of capitalism*, New York: Free Press