7 Telecommunications regulation

Andrea Renda

1. INTRODUCTION

Until the late 1980s, the telecommunications industry was dominated by fixed-line monopolists: many countries had a single network operator and provider of telecommunications services, most often owned by the state. These monopolists were typically subject to state regulation, similar to what occurs in other network industries: accordingly, they were subject to universal service obligations, which forced them to connect every citizen at affordable retail prices, regardless of the cost associated with reaching that customer. The two main models used to regulate telecommunications prices in the past decade have been Rate-of-Return (RoR) regulation, implemented for example in the United States for AT&T since the 1970s; and price-cap regulation, developed by Stephen Littlechild in the UK and applied to all British liberalizations of network industries. There are several differences between the two models, which can briefly be summarized as follows (see i.a. Johnson, 1989):

- Under a RoR regulation, monopoly firms are required to charge the price that would prevail in a competitive market, which is equal to efficient costs of production plus a market-determined rate of return on capital. RoR regulation has been criticized because it encourages cost-padding, and because, if the allowable rate is set too high, it encourages the adoption of an inefficiently high capital-labour ratio. This is called the Averch–Johnson effect (Averch and Johnson, 1962).
- Price cap regulation adjusts the operator’s prices according to the price cap index that reflects the overall rate of inflation in the economy, the ability of the operator to gain efficiencies relative to the average firm in the economy, and the inflation in the operator’s input prices relative to the average firm in the economy. Revenue cap regulation attempts to do the same thing, but for revenue rather than prices. Price cap regulation is sometimes called “CPI-X”, (in the United Kingdom “RPI-X”) after the basic formula employed to set price caps.1

---

1 CPI stands for Consumer Price Index; RPI for Revenue Price Index.
The debate on the most appropriate form of retail price regulation is, today, a historical matter rather than a contemporary one: other issues are more heavily debated, including net neutrality and interconnection models for the Next Generation Networks (NGN) era (see below, Section 4). As a matter of fact, the last three decades have witnessed sea changes in telecoms markets and, correspondingly, also in the focus of telecoms regulation. Suffice it to recall that today, thousands of fixed-line telecom operators are active around the world, countless mobile providers compete to gain new customers in developed and developing countries, and fixed-line, wireless, cable and satellite operators are increasingly thought to compete for the same customers. Many families are becoming “mobile only”. The advent of wireless telephony has revolutionized the way we live, and promises spectacular new developments from high-speed Internet surfing to mobile banking, to e-health services. Spectrum has consequently become one of the most precious assets in this policy domain, and the problem of allocating it efficiently is one of the key challenges of the next decade. At the same time, investment in fibre networks and the development of cloud computing are expected to bring massive benefits to citizens and businesses in the next few years: however, investment in fibre networks seems to proceed very slowly, also due to the economic crisis that has hit the global economy since 2007.

The next pages summarize the evolution of telecommunications regulation in the past three decades, and highlight the main policy issues that regulators around the world are trying to address today. The reader should be aware of one issue: the literature on telecommunications is vast and its scope quite broad; at the same time, there is little that law and economics scholars have contributed to the advancement of academic knowledge in the field, with some notable exceptions, such as Ronald Coase’s 1959 paper on the Federal Communications Commission. Accordingly, the next pages will have to depart from law and economics, including papers that belong more to the realm of industrial economics. At the same time, this chapter only reports the main conclusions to have emerged in the literature on a limited set of topics related to telecommunications regulation. For example, technical issues such as universal service, carrier selection/pre-selection, rights of way and other regulatory issues are not fully covered; on the contrary, more emphasis is devoted to unbundling and access policy, due to their relevance for mainstream law and economics and the extensive coverage in the scholarly literature. At the same time, key competition policy issues – for example, margin squeeze cases – are not covered in this chapter: even if the majority of margin squeeze cases were related to the telecommunications sector, margin squeeze can be considered as a general infringement of competition law, and as such should be kept outside the analysis of a particular sector. Finally, technical issues related to cloud computing and the Internet are not described in this chapter.
With these limitations, Section 2 below is dedicated to fixed-line telecommunications, whereas Section 3 discusses regulatory issues in the wireless sector. Section 4 outlines the key economics of modern broadband platforms, and highlights the policy problems that may have to be addressed in the years to come.

2. FIXED-LINE TELECOMMUNICATIONS: MAIN REGULATORY APPROACHES

In most countries that have undergone the liberalization of the telecommunications sector, the need to unbundle the incumbent’s network elements has been considered to be the main way to achieve entry of new players without requiring massive infrastructure investment by new entrants. This is certainly true for the United States since the AT&T breakup and, later, with the 1996 Telecommunications Act, which adopted unbundling as the main instrument of liberalization for telecommunications services. At the same time, this is true for Europe since the “Open Network Provisions” (ONP) era, as well as with the current regulatory framework for electronic communications. Similar approaches can be observed in Canada and Japan, as well as in many other countries (see Renda, 2010).

However, with the advent of broadband communications, the desirability of unbundling policies has been increasingly brought into question. The key moment in this respect was certainly the U-turn by the US Federal Communications Commission (FCC), which declared the failure of the “stepping stones” approach based on the provision of access to “Unbundled Network Elements” (UNE) and decided to lift regulatory obligations on incumbents, introducing regulatory holidays for broadband communications (see Gans and King, 2003; Renda, 2007). Since then, the possible trade-off between competition and investment in broadband networks has never left the spotlight in the debate on telecoms regulation. It is no surprise, then, that some commentators have observed that, given the similarity between the US “stepping stones” and the EU “investment ladder” approaches, the latter too should be reconsidered in the direction of a more lenient, and possibly less intrusive, telecommunications policy (see Bourreau et al., 2010).

2.1 Foundations of Unbundling Policy

Unbundling policy is considered, in and of itself, an exception to the general rule that those owning an asset should be entirely free to exclude others from its use – the so-called *jus excludendi omnes alios* that is well known to scholars in property (and Roman) law. For example, the US Supreme Court has
called the right to exclude others "one of the most essential sticks in the bundle of rights that are commonly characterized as property".\(^2\) The basic precondition of an unbundling rule that forces owners to share their asset with third parties is the existence of a superior public interest, which overlaps with the equally important goal of protecting private property rights and their role in incentivizing investment. Accordingly, whenever we observe unbundling in practice, there must be something that trumps a strong enforcement of property law: in the telecoms field, this something is represented by the need to force competition into previously monopolized markets. This is why unbundling is typically a liberalization tool. Other, similar cases of unbundling that are frequently observed are competition law cases, in which an undertaking that owns an asset that is impossible to replicate technically or economically (a so-called “essential facility”) is eventually forced by competition authorities to provide access to the asset, normally at fair, reasonable and non-discriminatory (FRAND) conditions.\(^3\)

From a law and economics perspective, unbundling can be seen as a residual remedy chosen by the legislator to protect a given legal entitlement. In particular, unbundling represents the conversion of a property rule (a right to exclude) into a liability rule (a right to compensation in case of third party access). In his elaboration of the original framework designed by Coase (1960) and Calabresi and Melamed (1972), Ayres (2005) has re-defined this type of liability rule as a call option, i.e. the possibility, for a third party, to purchase an entitlement at a regulated price without having to negotiate directly with the entitlement holder. This type of remedy is often considered to be preferable to a property rule whenever (i) it is impossible to ascertain with precision \textit{ex ante} what allocation of the entitlement maximizes allocative efficiency; and (ii) transaction costs associated with negotiation between the parties are considered to be substantial (see Kaplow and Shavell, 1996; and Ayres, 2005). In the case of telecoms liberalization, the latter condition is very likely to be met, since – absent a regulatory obligation – incumbents might pretend access prices that would make them at least indifferent between sharing their infrastructure or keeping it for themselves, and this would likely represent too high a price to pay for new entrants, since price would incorporate (at least) a monopoly profit. Since, following Ayres and Nalebuff (1997), the new entrant would have a rather poor BATNA ("best alternative to negotiated agreement"), the contractual balance would probably tilt in favour of the incumbent in this negotiation.


\(^3\) Impossibility in economic terms implies that, though in principle feasible, replication would entail prohibitively high costs.
Over time, the law and economics literature has adopted a mixed approach towards the choice between property and liability rules. As observed by Nicita and Rizzolli (2004), two main schools of thought have developed since Calabresi and Melamed’s initial contribution: one stressing the dominance of property rules (Epstein, 1997), and the other advocating the superior efficiency of liability rules, regardless of the dimension of transaction costs that need to be sustained in protecting rights (Kaplow and Shavell, 1996; Ayres and Talley, 1995). In particular, two contributions are worth mentioning in more detail – those of Kaplow and Shavell (1996) and Lucian Arye Bebchuk (2001).

Kaplow and Shavell (1996) introduce a clear distinction between takings and negative externalities as scenarios in which the trade-off between property and liability rules may emerge. The superiority of liability rules over property rules emerges in the latter case, given that these remedies minimize information costs for the judge. On the contrary, in the case of takings, property rules are superior to liability rules. The case of takings is obviously much closer to that of access policy in telecom regulation than the case of negative externalities.

Moreover, Bebchuk (2001) considers the trade-off between property and liability rules from an *ex ante* perspective, and concludes that the choice of the remedy significantly affects incentives to invest in a given activity. Generally, property rules emerge as superior tools for encouraging investment in the first place, whereas liability rules favour co-existing (or subsequent) uses over initial investment. Depending on the preferences and goals of the legislator, the right mix of incentives can be struck by using one of the rules or a combination thereof.

To sum up, the problem of unbundling can be framed as the conversion of a full-fledged property right into a right protected by means of a liability rule, where the option price is established by a third party (a sectoral regulator, or a competition authority). Based on the findings of three decades of debate on law and economics, it is fair to state that a liability rule may be comparably or preferable to a property rule from an *ex ante* perspective only when transaction costs are very high and the firm holding the entitlement is adequately compensated by the access price set. In what follows, we explore the practical underpinnings of the unbundling debate, and discuss the issue of the “right” price level in the context of unbundling.

### 2.1.1 From theory to practice: existing approaches to unbundling

The applicability of the law and economics debate on property and liability rules to the choice of whether to mandate the unbundling of incumbents’ network elements is limited since the former approach has been so far mostly confined to the case of property takings or, in particular, nuisance and negative externalities.
The difference, in the case of telecommunications as well as in antitrust law, is that in the latter domains we are not facing a conflict between property rights, but rather an overlap between conflicting public policy goals. More particularly, the interest in achieving and preserving competitive markets has been converted over time into regulatory prescriptions that carve in stone the relevance of liability rules as a tool for promoting competition against the undue monopolization of essential assets. Common carrier obligations in the US, as well as the essential facility doctrine (in many countries worldwide), are clear examples of this regulatory trend, especially in network industries, and particularly in telecommunications (Renda, 2010).4

A notable exception in the literature is Henry Ergas (2008), who analyses the law and economics behind Part IIIA of the Australian Trade Practices Act, a provision that allows for a generous application of the essential facilities doctrine in Australia, under the lens of Calabresi and Melamed (1972). He concludes that, before introducing a liability rule in a competition-oriented legislation, the following conditions must at least be met:

- The potential access provider must have control over an essential input, and use that control to extract monopoly rents in the downstream market.
- The refusal to supply access must be related to the protection of those rents by means of the exclusion of no-less-efficient competitors, rather than by the seemingly more plausible desire to preserve the efficiencies of vertical integration.
- It must be more efficient to regulate the monopolist by mandating the supply of access than by directly regulating the supply of the final goods that it monopolizes or simply not regulating at all.

Ergas then concludes that Part IIIA fails to properly test for these factors and creates a substantial risk of Type I errors, or “false condemnations” – that is, of mandating access when it should be denied.

Similar arguments can be developed for the essential facilities doctrine that has dominated the scene in telecommunications regulation in the United States, in Europe and in other countries. As recalled in Renda (2010), the doctrine was developed in the United States as early as the 1970s and was explicitly defined in a federal case regarding the telecoms sector (in the Verizon v. MCI case in 1993), but was also turned down explicitly by the Supreme Court in cases such as Trinko and LinkLine, where the Supreme Court declared to have never recognized the doctrine, which can at best be

---

4 Common carrier regulation refers to a situation in which an enterprise is obliged to carry anyone’s traffic over the resulting infrastructure, at regulated rates.
considered an elaboration of lower courts. Today, in the US the conditions for applying the essential facilities doctrine outside clearly defined regulatory provisions (such as those still in force for narrowband communications under Title II of the 1996 Telecommunications Act) are narrow at best. Likewise, in telecommunications regulation the FCC has initially adopted a rather broad interpretation of the essential facilities doctrine, especially with the 1999 Line Sharing Order, which assumed that “lack of access would materially raise the cost for competitive Local Exchange Carriers (LECs) to provide advanced services [such as DSL] to residential and small business users, delay broad facilities-based market entry and materially limit the scope and quality of competitor service offerings”. In 2002, the Line Sharing Order was vacated by the DC Circuit, which recognized that the existence of facilities-based competition (from less regulated cable operators) would not support mandatory unbundling of network elements as foreseen in the 1996 Telecommunications Act. This was simply the beginning of a declining trajectory, which led the essential facilities doctrine to be increasingly held as a residual province of telecoms regulation and antitrust in the US (see Renda 2010).

In Europe, the fate of unbundling and the essential facilities doctrine has been much less tragic. From a regulatory perspective, the entire regulatory framework for e-communications, from its first version in 1998 to its subsequent reforms in 2002 and 2009, is entirely based on the notion of unbundling and the promotion of service-based competition in the short term, as a way to stimulate long-run facilities-based competition. The idea of reconciling short- and long-term goals was brilliantly summarized in the concept of the “investment ladder”, which still represents the dominant model in telecom regulation across the EU27 (Cave, 2006b; Bourreau et al., 2010). It is fair to state that there could be hardly any better alternative to the investment ladder model in Europe, given the absence of facilities-based competition in many member states and, even more importantly, the need to achieve convergence in national regulatory approaches and to trigger market integration – something that would have been very difficult to achieve without a reference regulatory model. Evidence, in any event, testifies to a very mixed track record for this attempt to extend unbundling obligations over different rungs of a common


6 The ladder of investment theory postulates that the entry of players into the market should be achieved gradually, first allowing entrants to act as mere service providers, and later encouraging them to invest in network infrastructure.
ladder, especially when it comes to incentives to invest in new broadband infrastructure (see Bouckaert et al., 2010).

In Japan, Canada and many other countries around the world, the issue of unbundling and its controversial impact on incentives to invest has emerged in the past few years (Huigen and Cave, 2008): hence it seems fair to conclude that a trade-off between competition and investment exists in modern telecommunications networks. Recent literature has highlighted a mixed relationship between the extent of access policy and the level of investment in telecoms infrastructure. Research undertaken *inter alia* by Hausman and Sidak (2005); Oldale and Padilla (2004); Waverman et al. (2007); Wallsten and Hausladen (2010); Bourreau et al. (2010); Friederiszick et al. (2008); and Bouckaert et al. (2010) have shown rather mixed – if not discouraging – results from the implementation of the model in a number of countries. To the contrary, the recent study by the Berkman Center on Internet & Society at Harvard University reviews as many as 57 empirical papers and questions the reliability (and sometimes the independence) of their findings, arguing that open access policies have contributed to the success of the top performers in broadband penetration worldwide. Cambini and Jiang (2009) provide a very comprehensive survey of the theoretical and empirical literature on regulation and investment.

The academic debate is far from over, and recent authoritative opinions seem to point at a more balanced view of the interplay between competition and investment, with a growing role for non-regulatory public policy (as in Bauer, 2010). The emerging wisdom on how to strike the right balance between access policy and giving leeway to incumbents’ investment seems sophisticated enough to distinguish between geographical segmentation (even within the same country), technological conditions, and existing policies at all layers of the emerging Information and Communication Technology (ICT) infrastructure.

In a nutshell, whatever the solution to the investment ladder puzzle, the puzzle has changed, and unbundling seems to be even more difficult to implement in an NGN environment. Section 2.1.2 below comes back to this issue.

### 2.1.2 Unbundling, ladders and essential facilities

A quick reflection on the meaning of unbundling and its relationship with the essential facilities and the investment ladder is needed, in both legal and economic terms. In particular, it is important to recall that the long-standing jurisprudential elaboration that led to the establishment of a case for access to essential facilities (under fairly restrictive conditions, and not everywhere) in some countries does not necessarily support the legal and economic case for unbundling and an investment ladder approach.
For example, the European Court of Justice (ECJ) has clarified on several occasions the cumulative conditions that have to be met for compulsory third party access to be forced under community competition law in a stream of cases that goes from *Commercial Solvents* to *Tiércé Ladbroque, Bronner, Magill* and *IMS Health*. These conditions include, but are not limited to, the fact that the asset is indispensable for a third party to compete in the secondary market. In addition, Community competition law requires that the refusal to provide access “is preventing the emergence of a new product for which there is a potential consumers demand, that it is unjustified and such as to exclude any competition on a secondary market”.7

In another refusal to deal case decided by the ECJ, *Bronner*, Advocate General Jacobs observed that the application of any duty to deal must be very cautious, since making it easy for competitors to access the facilities of dominant firms would result in an absence of incentives for such competitors to develop their own facilities, while it would also remove incentives for dominant firms to invest in their own facilities – hence the reference to the special attention to be paid whenever the facilities are the “fruit of substantial investment”.8 This reference survives in the Commission Guidance document on the treatment of exclusionary abuses under article 82 EC Treaty (now 102 TFEU), published in December 2008. There, at §75, the Commission states that

The existence of … an obligation [to supply] – even for a fair remuneration – may undermine undertakings’ incentives to invest and innovate and, thereby, possibly harm consumers. The knowledge that they may have a duty to supply against their will may lead dominant undertakings – or undertakings who anticipate that they may become dominant – not to invest, or to invest less, in the activity in question. Also, competitors may be tempted to free ride on investments made by the dominant undertaking instead of investing themselves. Neither of these consequences would, in the long run, be in the interest of consumers.

The Commission also adds that “the termination of an existing supply arrangement is more likely to be found to be abusive than a de novo refusal to supply” (§84), and that “the Commission will consider claims by the dominant undertaking that a refusal to supply is necessary to allow the dominant undertaking to realise an adequate return on the investments required to develop its input business, thus generating incentives to continue to invest in the future, taking the risk of failed projects into account” (§89).

---

The above statements are interesting since they show one important feature of the EU version of unbundling and the investment ladder: that the conditions for mandating access to essential facilities under competition law are way more restrictive than those for gaining access to an incumbent’s telecoms infrastructure under the 2002 regulatory framework for e-communications. This is perfectly mirrored in two additional facts: (i) many of the relevant markets pre-defined by the European Commission in its 2003 Recommendation would not be defined as stand-alone markets by a competition authority; and (ii) the investment ladder approach by itself transcends the notion of essential facility and indispensability to build a broader framework for enabling gradual investment by new entrants. To put it roughly, the investment ladder is broader than the international notion of unbundling, and unbundling is broader than the essential facilities doctrine.

In addition, the emphasis placed on investment issues by the European Commission’s guidance paper seems to cast some doubts on the fact that the sectoral regulatory framework in the EU is really inspired by notions and tools borrowed from Community competition law. To be sure, it raises an important question: how do we deal with the unbundling of essential facilities, if the facilities in question have not been built yet? Section 4 below will return to this issue.

### 2.1.3 Setting the right access price

A discussion of the efficiency and desirability of mandatory access to essential facilities cannot ignore the issue of pricing: whether a liability rule can effectively balance investment incentives and the interest of competition – or better, pluralism – in a given market depends also on the price level that is set for access to the infrastructure. In this respect, regulators have generally opted in favour of forward-looking cost models such as Total Element Long Run Incremental Cost (TELRIC) or Total Service Long Run Incremental Cost (TSLRIC), rather than measures of opportunity cost or option value. These models are based on the notion of incremental costs, defined as “the total current output of that product” or – as defined by Panzar (1989) – “the change in the firm’s total cost caused by its introduction at the level \( y_i \), or, equivalently, the firm’s total cost of producing \( y \) minus what that cost would be if the production of good \( i \) were discontinued, leaving all other output levels unchanged.” See Gans and King (2003) for a comparison between the two models. The TELRIC model was used by the Federal Communication Commission in the United States under the 1996 Telecommunications Act, while the TSLRIC model, renamed Long Run Average Incremental Cost (LRAIC), is the reference model used in the European Union. The need to take into account common costs – normally not covered by the concept of incremental costs, which are product-specific – has led regulators to develop over
time slightly different models, such as TSLRIC+ (which includes indirect costs) and TSLRIC++ (which accounts for the so-called “access deficit contribution”).

In any event, such cost measures are often considered to systematically under-estimate the remuneration level needed to compensate the economic operator for its investment in infrastructure, since “to persuade a [competing provider] to invest, the access price must cover both the competitor’s cost of supply and the value of the option that the investment would destroy. If the option is not priced in the access charge the competitor’s incentives will be distorted against investment” (Cave, 2010). A related issue is whether the access price should mimic a perfectly competitive price, or should indeed attempt to replicate the price that would result from a negotiation between the incumbent and the new entrant, if transaction costs were not prohibitive. Needless to say, the former option corresponds to the current regulatory practice worldwide; however, a law and economics approach to the issue would suggest that, if the real purpose is to replace a property rule with a liability rule to avoid transaction costs, then there is no need to combine a taking (the compulsory access) with allocative inefficiency (the under-remuneration of the value of access). If the “bundle of rights” that composes property is to be limited with the imposition of compulsory access, remuneration of that access should not be purely cost-based (plus, maybe, WACC), but should include the surplus the incumbent would have derived from negotiation with an (equally efficient) entrant. That said, Pindyck (2005) observed that “inordinate time and resources have been spent on the question of whether an [incumbent’s] historic average cost of capital is 12.9 versus 13.1 percent. We have seen that the correct cost of capital input for TELRIC should be significantly larger than either of these numbers; at issue is whether it should be larger by about 1 percentage point or 4 or 5 percentage points.”

The issue of risky investment has recently resurfaced in the debate over the right price for access to next generation access networks. The idea of adding a risk premium to the usual formula for access pricing to remunerate investment in high-speed networks was launched by the European Commission in its 2010 Next Generation Access (NGA) recommendation. However, the additional risk has never been fully demonstrated, and its logical and economic foundations appear rather shaky, or – better – a way to preserve the overall EU approach to access policy, but at the same time showing some consideration for incumbents’ incentives to invest.

9 The access deficit contribution is the shortfall between the cost of providing basic access and the revenues that the incumbent operator is able to secure under the price control regulations.
As a result, the case for unbundling of incumbent networks is still very controversial in the literature, including the empirical research. The dispute over the “right” formula to calculate access prices, together with mixed evidence on the impact of access policy on market performance, are very likely to continue in the years to come. This is potentially even more troublesome because, while academics and practitioners try to square the circle by setting a regulatory approach that reconciles investment and competition, the dynamics of convergence are posing many more fundamental questions to the policymakers. For example, are we sure that incumbents’ networks should be considered as essential facilities today? Are we sure that a player with a substantial share of the fixed-line (sub-)market would be able to exploit any market power and extract any rent from its position? If these two are considered to be preconditions of access policy, addressing these questions is now of utmost importance for the future regulatory approach in the era of high-speed telecommunications networks. Section 4 below returns to this fundamental issue.

3. WIRELESS TELECOMMUNICATIONS

The rationale that has led to the liberalization of fixed-line telecommunications cannot be applied to wireless ones, which exhibit totally different features, in primis the absence of a legacy monopoly. The possibility of transmitting information without wires, by using radio frequencies available in the ether, has become known since Guglielmo Marconi’s experiment with the transmission of the three-dot Morse code for the letter “S” over a distance of three kilometres in 1895. Today, more than a century after that experiment, wireless communications have changed the way we live. Among the most prominent forms of wireless communications – together with broadcasting, amateur radio, satellite services, military uses etc. – is the mobile transmission of voice and data, which forms the basis for modern wireless telecommunications. In 2002, with one billion users worldwide, mobile communications for the first time surpassed fixed-line subscribers (ITU, 2003; Garbacz and Thompson, 2007). In 2008, the International Telecommunications Union (ITU) estimated that there were over 4 billion mobile users and approximately 1.2 billion fixed lines. That is, the future is mobile, and even more so in developing countries, where fixed line does not have high penetration rates.

Gans, King and Wright (2005) divide the history of mobile communications into four main periods: (i) a “pre- cellular” period that involved mobile telephones that exclusively used a frequency band in a particular area, such as German A-Netz and B-Netz telephones; (ii) first generation cellular mobile telephones developed in the 1980s such as the Advanced Mobile Phone
System (AMPS) in the US, the Total Access Communications System (TACS) in the UK, the C-Netz in Germany, the Nordic Mobile Telephone (NMT) in Scandinavia; (iii) the so-called second generation (2G) mobile telephones that used digital technology, including Global System for Mobile communication and other, incompatible technologies such as Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) technologies; and (iv) modern third generation (3G) phones based technologies such as the Universal Mobile Telecommunications System (UMTS), which have led to the emergence of what we call “smartphones” in today’s jargon. Today, compared to what Gans, King and Wright could describe in 2005, a new generation of cellular phones is being launched, with integrated fixed wireless and mobile technologies, which can in principle enable Internet surfing at speeds comparable to those available on a VDSL or a fibre connection. These fourth generation (4G) technologies include, most prominently, the Long Term Evolution (LTE) technology that is backward compatible with the UMTS.

Besides the description of the most widespread technologies in mobile communications, what is important from a law and economics perspective is that these devices, which initially enabled users to merely place voice calls over a mobile network, can increasingly replace a personal computer for the purpose of navigating the Internet, and support applications and services just as fixed-line broadband platforms can. As will be explained below, this “platformization” of the mobile business (Ballon, 2009) leads to very interesting consequences in terms of the dynamics of competition and the underlying economics. Today’s battle between Android-enabled smartphones and Apple’s family of iPod, iPhones and iPads is the ultimate result of a market that has been revolutionized by the availability of sufficient capacity to transmit data, besides voice.

Section 3.1 below illustrates the economic features of modern mobile platforms, including two-sided market models, the waterbed effect and the economics of handset subsidies. Section 3.1.1 is dedicated to the main pricing models and regulatory tools used in mobile markets: in particular, the section distinguishes between calling-party-pays and mobile-party-pays models and regulatory models for termination rates. Section 3.2 contains a brief illustration of the main contributions in the literature on spectrum management and allocation from a law and economics perspective.

### 3.1 Some Economics of Wireless Platforms: Payment Regimes, Two-Sided Markets and the Waterbed Effect

Mobile markets typically feature high fixed costs and low marginal costs

---

10 VDSL stands for Very-high-bitrate Digital Subscriber Line.
compared to traditional markets (Haucap, 2003). In these markets, the scarcity of spectrum and the high level of fixed costs constitute important entry barriers and lead to oligopolistic structures, which are often associated with intense competition in the market (Deventer and Haucap, 2005; Valletti, 2006). In addition, when mobile operators are asymmetric and able to discriminate between on-net and off-net calls, subscribers to a large network may face important costs that would discourage them from switching to a competing network (Lopez and Rey, 2009).

A first, very important determinant of the patterns of competition in mobile markets is the payment regime chosen. Countries around the world have experimented with two main models: the calling-party-pays (CPP) and the receiving-party-pays (RPP) models. Under RPP regimes, terminating mobile networks charge their own customers for termination services, whereas under CPP terminating operators charge the originating network. The CPP regime, thus, can lead to market power in termination markets and is widely accepted as justification for regulatory interventions (see Armstrong, 1998; Laffont, Rey and Tirole, 1998). On the other hand, following, for example, DeGraba (2000), Quigley and Vogelsang (2003), Valletti and Houpis (2005), Berger (2005) and Littlechild (2006), under the RPP regime mobile network operators have no incentives for charging monopolistic termination rates (but contra, see Gans and King, 2000; and Wright, 2002).

CPP is currently the most frequently used model around the world and in particular in the European Union, whereas in the United States the RPP regime has been adopted. In Europe, every national regulator has ended up regulating termination rates, allowing for total cost recovery based on fully allocated cost (FAC) models. This approach has been increasingly challenged by the two-sided market literature, which emphasizes the role that call externalities play in the analysis of competition, equilibrium pricing and entry in these markets (see Harbord and Pagnozzi 2010, and infra on two-sided markets). As reported by Harbord and Hoernig (2010), impetus for change has also come from the entry of new mobile network operators in many European countries. These operators argue that their growth and profitability have been hampered by high termination rates and by the significant levels of on-net/off-net price discrimination adopted by incumbent mobile network operators.

In the two-sided market literature, providers of mobile communications are described as multi-product platforms that need to efficiently match callers and receivers. Among the different services provided by mobile phone operators are the wholesale and retail origination of calls, the wholesale and retail termination of calls, the lease or sale of a handset, and roaming services. Within retail origination and termination markets, calls are further differentiated into on-net (between users that are both subscribers of the same provider), off-net (between subscribers of different operators) and calls to and from fixed
networks. A two-sided platform having to ensure an appropriate balance between callers and receivers will choose whether and where to obtain revenues by exploiting network externalities and increasing switching costs to the extent that this is possible.\[11\] This means that, depending on the market conditions and the characteristics of user demand, mobile platform operators will set prices in a way that is normally unrelated to the underlying cost of the service.

Theoretical works that address the issue include Armstrong (1997, 2002), Gans and King (2000), Thompson et al. (2007), Valletti and Houpis (2005), Valletti (2006), Wright (1999, 2002). This body of work broadly concludes that even highly competitive mobile operators will set termination charges to maximize the monopoly profit from termination and thereby the subsidy they can offer to their subscribers. Accordingly, the socially optimal termination charge is always above the cost level. Hausman and Wright (2006) improve on this work by allowing for the possibility that fixed-line callers may also themselves be mobile subscribers. Their additional conclusions are that mobile operators will set the termination charge below the monopoly level; and that equilibrium termination charges are not necessarily too high.

This is, after all, a corollary of a well-known effect in mobile networks, known as the “waterbed effect”, i.e. the effect whereby regulation of one of the prices of a multiproduct firm causes one or more of its other unregulated prices to change as a result of the firm’s profit-maximizing behavior. The term “waterbed effect” was used for the first time in 1997 during an investigation by the British Monopolies and Mergers Commission, and was extensively recalled in the debate over the regulation of mobile termination rates and also during the recent debate on the regulation of wholesale and retail roaming charges in the European Union since 2007. Genakos and Valletti (2008) provide a survey of the literature and empirical data on the existence of the waterbed effect in the EU mobile industry, which they find to be strong.

The combined effect of network externalities, two-sided market features, and the waterbed effect lead to a very uneasy case for traditional price regulation in the mobile sector. Nevertheless, in countries with CPP regimes, mobile operators have been found to be monopolists in the termination of calls on their own networks, thus deserving cost-based regulation of termination rates. The contrast between the findings of the literature and current regulatory patterns has not been solved to date, although the advent of smartphones and modern mobile platforms seems to have partly overcome this issue.

3.1.1 Facilitating entry in mobile communications

Regulators in the mobile sector have tried to overcome the barriers to entry identified in the previous section – availability of spectrum, high fixed costs, switching costs – by using a limited set of recurring tools, which feature mobile number portability and the introduction of mobile virtual network operators.

First, regulators have used Mobile Number Portability (MNP) to ensure that users could change providers without losing their number. The first formal analysis of MNP is due to Aoki and Small (1999), who examine the welfare effects of MNP for different levels of mobile penetration (or market saturation), finding that the overall welfare effect of MNP is ambiguous if the investment costs of implementing a MNP system are weighed against the benefits of more intense competition between mobile carriers. In related papers, Gans, King and Woodbridge (2001) and Haucap (2003) have focused on the question of how to allocate the property rights in phone numbers and the costs of implementing number portability. Bühler and Haucap (2004) and especially, Bühler, Dewenter and Haucap (2006) propounded that the static effects of MNP are on retail prices, price elasticities, termination charges and market shares, whereas the dynamic effects of MNP are on firms’ entry and investment. Shi, Chiang and Rhee (2006) show that MNP led to a higher market concentration due to on-net pricing in Hong Kong.

Second, regulators have sought to stimulate the entry of new mobile players despite the absence of awarded spectrum by introducing Mobile Virtual Network Operators (MVNO), which provide mobile phone services but do not have their own licensed frequency allocation of radio spectrum, nor necessarily have all of the infrastructure required to provide a mobile telephone service. MVNOs are thus pure resellers of the mobile service, although in some cases they also own some infrastructure. For a taxonomy of MVNOs and their impact on the market, see among others Brito and Pereira (2008), Ordover and Shaffer (2007) and Kalmus and Wiethaus (2010). Overall, the conclusions of this literature point to the mixed effects of the entry of MVNOs in the short run, and negative effects in terms of investment and dynamic competition in the long run, very much in line with the findings of the literature on access policy in fixed-line communications.

3.2 Spectrum Management and Allocation

Since the birth of modern communications, spectrum policy has always implied the top-down allocation of licences by administrative authorities for the provision of specific services at specific frequencies. Such a “command and control” policy approach was chiefly dictated by the need to avoid interference between conflicting uses. In this respect, use of the spectrum, especially at certain frequencies, can create situations of “nuisance” very similar to
those that arise in property law. Based on these potential interference problems, until the 1950s there was very little debate on the ways in which spectrum could be allocated. Given the relative under-development of wireless communications and the existence of (mostly state-owned) monopolies in telecommunications and audiovisual services, regulators simply decided “who” should use “which” portion of the spectrum, to deliver “what” services, and “how” (i.e. with what technology).

During those years Leo Herzel (1951) started questioning this approach. In 1959, Ronald Coase wrote an important article, simply entitled “The Federal Communications Commission”, where he argued in favour of a market for trading property rights on the radio spectrum. Compared to a scenario in which the state allocates spectrum rights with a top-down, command and control procedure, Coase argued that market forces can perform better due to superior information. Since the government is not expected to have enough information to understand who will be able to use each portion of the spectrum most efficiently, and how this changes over time, reliance on market forces to achieve allocative efficiency would have been the best way to unleash the extraordinary economic potential of the airwaves. As is well known, Coase’s argument was not welcomed by regulators. Asked to testify at a hearing, he was greeted by a member of the Commission: “Good morning, Professor Coase. Please tell us, is this all a big joke?” Today, spectrum markets are at the core of the policy agenda in at least some industrialized countries (e.g. the US, UK, EU).

Some European countries, such as Germany and the UK have advanced on market-based mechanisms in spectrum, using spectrum auctions since the late 1990s, following Coase (1959) and later developments (Benzoni and Kelman, 1993). Trading some of the radio spectrum in specific bands like a commodity is currently permitted only in a few countries. An alternative and more radical approach is to permit unlicensed spectrum bands for anyone to use – the basis of the “commons” in spectrum. This concept and its variants such as the “supercommons” hold that spectrum is not a commodity, and is certainly not scarce, but rather it is just misallocated (Werbach, 2004). So regulatory proposals based on the spectrum being a physical asset, defined in bands of frequencies, artificially constrain exploitation as a common good. Extensive use of a commons approach depends on new radio technology to enable far more sharing of the spectrum and so refocusing radio regulation away from the spectrum itself and towards the devices used for communication – the position of the computer industry rather than the telecommunications sector. New technologies that promise a smarter and more agile use of the spectrum include Software Defined Radio/Cognitive Radio (SDR/CR); spatial multiplexing using Multiple-input, Multiple-output (MIMO) systems; mesh networks; spread spectrum; compression; bit rate encoding and others. For an explanation, see Bohlin et al. (2008).
More generally, the literature has focused on the relative advantages of three alternative modes of spectrum allocation: command and control, market-based mechanisms, and the commons. For a survey of the pros and cons of the various models, see Faulhaber (2005, 2006) and Bohlin et al. (2008). Generally speaking, there is no one-size-fits-all answer to the question which model is the most appropriate. Different portions of spectrum feature different characteristics in terms of coverage and capacity. Accordingly, some portions of spectrum are more exposed to interference than others, and some portions are more suited to certain technologies than others. This issue has become even more evident in the current debate on the digital dividend. The coordination of spectrum at a global level is the responsibility of the International Telecommunication Union (ITU), a United Nations agency with the mission to maintain and extend international cooperation for the improvement and rational use of telecommunications. Every three to four years, the ITU-R holds a World Administrative Radio-communication Conference (WARC, now abbreviated to WRC), a process aimed at adapting the ITU Radio Regulations (RR), the international treaty coordinating spectrum usage globally. At WRCs, frequencies are first allocated to services (referred to as allocations); subsequently, individual countries allot frequencies to specific areas or regions (referred to as allotments) and assign them through licenses to stations (in a process called assignment). The allocation of each band within the regions may be to one or more services within one of the two categories: primary services or secondary services.

Besides spectrum management and allocation modes, the initial award of spectrum has also been subject to extensive literature, in particular after the (in)famous 3G auctions that led to the award of spectrum for UMTS cellular phones. Failures in auction design and in setting the overall objectives of the auctions – mostly focused on maximizing government revenues, to the detriment of operators and ultimately of consumers – have been highlighted by key auction experts such as Peter Cramton (1995, 2001, 2002), Daniel Sokol (2001), Paul Klemperer (2002, 2004) and Tom Hazlett and Roberto Muñoz.

---

12 During 2000 and 2001, several European governments used auctions to award licences to use spectrum bands for the provision of third generation wireless telephony services (3G). In most countries, auctions were designed mostly to maximize government revenues, rather than consumer welfare. This led to huge costs for mobile operators, which generated undesirable consequences for end users, in terms of both reduced investment in network deployment, and higher retail prices. As auction theorist Paul Klemperer clarified in 2002, there were enormous differences in the revenues from the European 3G auctions, ranging from €20 per capita in Switzerland to €650 per capita in the UK, though the values of the licences sold were similar. In addition, “poor” auction designs in some countries facilitated collusion between firms and failed to attract entrants.
That said, auctions are often considered to be superior to alternative award modes, such as direct award or beauty contest. For example, Prat and Valletti (2001) observe that the widely used Simultaneous Ascending Auction has many advantages, but may also facilitate collusion and discourage bids on more than one licence.

Spectrum auctions for 3G have been run by several countries around the world, including, recently, India and Turkey. The rather discouraging results achieved with 3G auctions have not stopped national government from pursuing the award of further portions of spectrum through auction mechanisms. Notable recent examples include auctions for the award of spectrum in the 700 MHz band in the US and in the 800 MHz band in many European countries for the award of the so-called “digital dividend” from wireless broadband technologies. The digital dividend is the portion of spectrum that can be potentially reallocated due to the changeover from analogue to digital TV. In particular, channels 61–69 or the spectrum portion between 780 MHz and 860 MHz are considered to be ideal for deploying 4G mobile networks without facing excessive costs. The European Commission is currently proposing a EU-wide coordinated strategy to award this portion of the digital dividend to mobile operators for the deployment of 4G technologies.

### 4. ACCESS POLICY AND OTHER REGULATORY ISSUES IN NEXT GENERATION NETWORKS

Many of the traditional paradigms in telecommunications regulation are being revolutionized by the transition towards IP-based infrastructure. Heavily debated issues include: (i) the scope of universal service – in particular, whether it should be expanded to include high-speed broadband connections and mobile telephony (see Blackman and Forge, 2008; Bohlin and Teppayayon, 2009); (ii) geographical segmentation – i.e. whether market analysis and corresponding remedies should be made more dependent on the peculiarities of the territory, in particular as regards the business case for deploying competing broadband networks (Xavier, 2010); (iii) the scope of future remedies, such as duct-sharing, in-building wiring, sub-loop unbundling, etc.; (iv) the future of access policy and its impact on incentives to invest (Huigen and Cave, 2008).

A heavily debated issue is the type of interconnection regime that best suits the NGN model. In particular, while “Calling Party’s Network Pays” (CPNP) regimes have been the most widely used in telecommunications, the Internet has always been essentially working on a “bill and keep” (BaK) model. Economists such as DeGraba (2000); Littechild (2006); Marcus (2008) and Harbord and Pagnozzi (2008) have compared various interconnection...
regimes, finding that: (i) the transition towards NGNs will lead prices per minute to fall, narrowing down the differences between a CPNP and a BaK regime; (ii) BaK can reduce regulatory cost and uncertainty and increases incentives for cost minimization as more costs are subjected to competitive cost recovery; (iii) BaK internalizes call and network externalities better than CPNP; (iv) BaK is expected to lead to higher average usage per capita and a lower average price per minute; (v) BaK could possibly lead to a slightly lower handset ownership; (vi) BaK is likely to deliver a material welfare gain to consumers overall, whereas it will exert mixed effects on operators, leading to an adjustment of the competitive balance between fixed and mobile operators. More neutral impacts are expected on cost efficiency and on Quality of Service. This view, shared also by the European Regulators Group (now BEREC) in 2009, is still challenged by many mobile operators, and as such cannot be considered conclusive.

Even more importantly, the case for network unbundling as the core remedy to promote sustainable and vibrant competition in modern communication networks must today be considered against developments in competitive dynamics triggered by several dimensions of convergence: not only convergence (and increased competition) between fixed and mobile communications, but also convergence between the telecommunications and the IT and media domains, and convergence between the infrastructure layer and higher layers of all-IP networks. In more detail:

- **Convergence between fixed and mobile telecommunications is finally becoming a reality.** This is certainly happening, though slowly, in Europe, as confirmed by a recent decision adopted by the European Commission, which authorized the definition of a common fixed-mobile relevant market for retail broadband in Austria. The Commission recalled that “[…] fixed and mobile retail broadband services are normally not belonging to the same market. However, on the basis of the following circumstances closely related to the specificity of the Austrian market, the Commission accepts the inclusion of mobile and broadband connections into the retail residential market for the purposes of the present notification.” Further prospects in this direction came from a recent document jointly elaborated by the BEREC and the Radio Spectrum Policy Group (2010), which discusses the main conditions for defining joint fixed-mobile markets. The use of femtocells13 and the

---

13 Femtocells are small cellular base stations, typically designed for use in the home or a small business. They connect to the service provider’s network via broadband (such as DSL or cable), allowing mobile operators to improve coverage and
remarkable speed of imminent 4G networks suggest that the substitutability between fixed and mobile broadband access will be on the increase in the months to come.

- Convergence between telecommunications and IT is fully realized by the migration towards an all-IP infrastructure, which is bringing new business models, the creation of multi-layered platforms where applications and services dominate user experience, and constantly changing competitive dynamics. Not only are fixed broadband platforms increasingly integrated into the Internet, but cloud computing is shifting most of the computing capacity into centralized servers, which will be made accessible from both fixed and mobile devices. The success of the App stores created by Apple and Google Android also promises to revolutionize the way in which we use computers, not only smartphones. This form of convergence is also triggering convergence between the infrastructure layer and higher layers of all-IP architectures, such as the logical layer, the application layer and the content layer in the (simplified) OSI representation (see Figure 7.1).

As stated, among others in OECD (2009), broadband platforms are much more than simple communications networks, and can be considered as ecosystems that comprise “different elements that use high-speed connectivity to interact in different ways”. In these ecosystems, competitive dynamics have become far more complex than used to be the case when the telecoms sector resembled a traditional network industry, mostly posing problems of third-party access and liberalization. The foundations of unbundling policy become even shakier when we look at the features of emerging markets, for the following reasons.

First, the emerging substitutability between fixed and mobile has a direct effect on the nature of essential facilities often attached to the incumbent’s fixed network. Even when reasonably substitutable fixed networks are not available, the existence of wireless solutions that fall in the same relevant market clashes with one of the conditions for a finding of essential facilities, i.e. the impossibility of technically or economically replicating the service. Absent this lack of replicability, unbundling seems to be much less justified.

capacity, especially indoors, achieving a quality of service and speed similar to those of the fixed-line telecoms networks.

14 In non-technical terms, cloud computing can be defined as location-independent computing. It implies that end users’ applications and contents are stored in a data centre, and the user can access them remotely from anywhere. Of course, it requires an always-on connection to the internet; at the same time, it potentially generates enormous savings in terms of IT equipment and software, which can be “rented” remotely by business or private users without upfront fixed costs of acquisitions, and only for the time needed.
Second, the assessment of market power is becoming increasingly complex due to (i) “horizontal” competition coming from players that operate in the same relevant market as the fixed-line incumbents (facilities-based cable or fibre entrants, wireless broadband operators, consortia of municipalities, etc.); (ii) “vertical”, “intra-platform” competitive pressure exerted by players that provide competing services in a nomadic way (e.g. Skype or Google voice for VoIP services); and “inter-platform” competition by players that propose themselves as platform operators, even if they come from different relevant markets (e.g. Apple’s iPhone or iPad, Google Android, Nokia Ovi, and many other nascent platforms). The literature has summarized these dynamics of competition – and especially the latter one – by referring to “competition for eyeballs”, which is animated by competing platforms that try to conquer the attention (and the bill) of the end user. Cloud computing can do nothing but exacerbate this form of competition, with several private cloud managers offering closed, semi-open or fully open cloud services. The law and economics literature has not approached this issue to date, but contributions are expected to proliferate in the years to come.

Third, a related, procedural problem for regulators and competition authorities is how to define the relevant market. The links between system layers and the lack of fully interoperable standards creates hidden provinces in cyberspace, where substitutability between platforms or platform “complementors” is indeed limited, warranting narrow market definitions. Antitrust authorities have already found their way into this quagmire. For example, in the US Microsoft case the relevant market for Intel-compatible Operating Systems (OS) was considered to be separate from the relevant market for Mac-compatible OS. The FTC went even further in a famous case, *Intel v. Intergraph*, by defining Intel as a monopolist for Intel processors, something that should have at least rung a bell. The fact that in the ICT world, “the
license is the product” (Gomulkiewicz, 1998), and “the product can become the market” (due to network externalities and tipping, see Rohlfs, 1974; Arthur, 1989; Katz and Shapiro, 1985; Shapiro and Varian, 1999) suggests that the notion of relevant market, interpreted in the way we have done outside the ICT world, may become completely useless in modern broadband platforms.

Fourth, as recalled above for wireless platforms, it is now widely acknowledged that modern broadband platforms exhibit the features of two-sided, or more accurately, multi-sided markets (see i.a. Rochet and Tirole, 2003, 2006; Evans and Schmalensee, 2007, 2008; Gawer, 2009). No player can succeed in capturing the attention of new users in those markets without good network connectivity, wide participation by application and content providers, one or more compatible device producers, and of course an established population of users (Poel et al., 2009; Renda, 2010). This peculiarity creates, among other things, also problems in terms of the selection of appropriate remedies. In particular, cost-based pricing is in most cases inappropriate for these types of markets (Wright, 2003), and even asymmetric regulation — i.e. imposing stricter regulatory obligations only on some market players — can create problems, since behaviours that may be erroneously considered monopolization strategies are in fact replicated by all players in the market, regardless of their market power.

Where does this leave unbundling practices? The theoretical foundations of unbundling, as described in Section 2.1.1 above, are likely to be severely jeopardized by existing developments. In particular, policymakers will be forced to identify those elements of modern broadband platforms, the replication of which would be absolutely uneconomical, such that mandatory access is the most appropriate pro-competitive remedy.

As a matter of fact, for the infrastructure layer these elements seem to be heavily dependent on the “where” (geographic area), the “what” (some technologies are much more difficult than others when it comes to unbundling, e.g. GPON) and the “how” (how to arrange migration to the new ladder for LLU operators, whether to opt for access to in-house wiring, wavelength unbundling at the ODF, access to ducts, dark fibre, etc.). As of today, elements that may be difficult to replicate certainly include passive infrastructure (ducts, masts) and — under more restrictive circumstances — bit-stream or sub-loops. However:

- This reasoning is valid only in “1.x” regions, i.e. areas where there is only one fixed-line broadband network, together with wireless (up to 3G). With more facilities-based competition, replicability is already proven in practice, and the economic justification for unbundling is much weaker.
• Other equally important bottlenecks may be found in other layers – for example, the operating system; the DRM system; killer apps; privileged/discriminatory access to a dominant cloud; key content; billing/charging functions and even IPR-protected business methods can be seen as candidates for mandatory access policy.
• An additional problem, which is very often underrated or ignored, is that when we discuss essential facilities in regulation or competition policy, we are normally talking about something that is already in place – be that a press distribution system (Bronner), an operating system (Microsoft) or even a ski resort’s facilities (Aspen Skiing). Here, we are attaching the essentiality label to facilities that have to be built – no surprise that the competition-investment trade-off becomes even more urgent. Against this background, the choice of the appropriate remedy in the NGN environment resembles the ex ante framework adopted by Lucian Ayre Bebchuk (2001) rather more than the ex post one of Kaplow and Shavell (1996). Hence it is no surprise if, in countries where unbundling is likely to be on the horizon, incumbents have just decided to stay away from investing in NGNs.

4.1 The Ladder of Investment in the NGN Environment

The fact that unbundling practices must change substantially in an NGN environment is practically uncontroversial, and has been confirmed by several regulators and field experts in the past years (Cave, 2010; Huigen and Cave, 2008; BEREC, 2010). The main differences that are likely to emerge concerning the application of the ladder of investment are the following.

First, the ladder of investment is different compared with copper networks. Access points and conditions of replicability change dramatically from copper to all-IP networks. As explained in a recent note by the EU BEREC, and exemplified in Figure 7.2, both the access products and wholesale products available to reach the access point change significantly.

Moreover, the functioning of the ladder depends on the type of network and the specific technology used. For example, in an FTTC network there is much less space to co-locate equipment and far fewer premises connected to each site compared with traditional networks, since passive access can take place at the street cabinet only. A study by the UK regulator Ofcom has found that sub-loop unbundling for an FTTC network would increase the cost of provision by a minimum of 34%, rising to 37% in the case of three additional providers.15 On the other hand, physical unbundling for a passive

15 See Ofcom (2010).
optical network (PON) is hardly practical, though it could theoretically occur at the splitter level. The easiest case for unbundling can be made for the most expensive networks, i.e. peer-to-peer fibre-to-the-home (p2p FTTH) networks: however, given that the investments required are very substantial, one may end up questioning the opportunity to mandate access to those networks.

Against this background, the emerging approaches in countries that are implementing access policy for Next Generation Access Networks (NGANs) tend to focus mostly on the sharing of passive infrastructure, and in particular duct-sharing, rather than on active infrastructure sharing (such as bitstream or sub-loop unbundling, SLU). The scope and conditions for infrastructure sharing, therefore, change significantly, together with the conditions for effective competition between incumbents and new entrants. In other words, whether the investment ladder can be as effective in an NGAN environment, as it has proven to be in traditional copper networks, is unclear at best.

In addition, and also as a consequence of the array of available technologies, next generation access networks call for geographical segmentation in

Source: BEREC (2010).

Figure 7.2 The “new” ladder
the identification and application of regulatory remedies. This increases complexity in the regulator’s everyday activities, as it calls for customized remedies for every portion of territory. As acknowledged by the European Commission (2010),

The deployment of fibre networks is likely to modify the current network topology and access points (in particular in relation to LLU), thus affecting the investments made. It is necessary that NRAs adopt a proactive regulatory approach which promotes investment by the incumbent and alternative operators, whilst preserving the investments already made by alternative operators in [local loop unbundling].

This also means that the challenge for policymakers has now become essentially fourfold, as they must seek to:

- **Preserve the incumbent’s incentive to invest.** Deployment of high-speed broadband networks is considered to provide a beneficial boost to the economy in terms of growth, jobs and productivity. The goal of stimulating investment has become even more important in recent times, as counter-cyclical investment in broadband networks was evoked in several countries – plus, international competition to rank high in broadband deployment has become frantic.

- **Preserve the incentives of those that have already purchased LLU.** Players that have made their way into the incumbent’s copper network by purchasing access to unbundled local loop may find it very difficult to jump to different rungs of a different ladder, given the significant size of the investment already undertaken (see Cave, 2010; Bourreau et al., 2010).

- **Preserve the incentive and viability of “new new entrants”**. Devising a pricing policy aimed at providing incentives to current LLU holders to migrate to the next generation access network is not the same thing as providing incentives for brand-new entrants to climb the investment ladder from scratch. This may create substantial problems for regulators in the first years of transition towards new all-IP networks.

- **Keep prices down for end consumers.** At the end of the day, policymakers also have to ensure that whatever pricing strategy is in place, end prices for consumers are affordable, so that the demand for NGN subscription remains sufficiently high.

Whether policymakers wishing to embark on this endeavour will be able to strike the balance between these four objectives is matter for future evaluation. To be sure, the future of unbundling crucially depends on the tools that will be utilized in order to achieve a mix of these four results.
4.2 Net Neutrality and the Telecoms-IT Interface

As telecommunications networks developed the capacity to carry digitized data at high speed – the core network in most industrialized countries is already made of fibre – retail broadband access has also become the key toll to enter the Internet. Inevitably, this has potentially left telecom network operators – especially fixed-line ones at the outset – with some degree of control over users’ bills and behaviour. In 2005, a small telecom operator named Madison River decided to use this degree of control to prevent its subscribers choosing a competing VoIP provider (Vonage) once on the Internet. The famous Madison River case ended with a negligible fine, giving rise to the most furious debate ever seen in the telecoms world, especially in the US, where after half a decade, it is still far from its final word.

Arguments in favour of regulatory intervention to mandate full net neutrality and keeping telecom networks as “dump pipes” developed with exclusive reference to the infrastructure and logical layers of the value chain. On the one hand, telecom operators claimed that the impossibility of managing traffic on their networks would have jeopardized the quality of the user experience, denied the possibility of a more efficient and effective provision of the Internet service, and left the whole Web prey to spam and illegal p2p file sharing, which – despite its illegality – continued for a long time to represent roughly half of all Internet traffic. On the other hand, “neutralists” challenged this view by stating that the end-to-end nature of the Internet should not be contaminated by intelligence at the core of the network, which would reduce the value of the network due to filtering of content and speech and the narrowing down of spaces for creativity at the edges.

Today, the debate seems to have evolved towards the recognition of the importance of traffic management for certain, specific purposes (e.g. spam filtering), coupled with defining those traffic management practices that can be considered reasonable, and under what circumstances. Most notably, there was a strong reaction by academics to the notice of proposed rulemaking (NPRM) published by the FCC in late 2009, announcing the intention to regulate ISPs’ behaviour to ensure the neutrality of the network, with the exception of yet-to-be-defined reasonable traffic management practices. As already mentioned in the introduction to this chapter, the debate has also raged in the European Union, both in the Commission and in the European Parliament.

Several academics have gone back to the theory of regulation and the particular economics of the Internet ecosystem to assess the soundness of policies studied around the world. For example, a sizable group of academics from several parts of the world filed a submission with the FCC to state that, in their opinion, the NPRM is not grounded in economics since it fails to demonstrate the existence of a market failure (Brito et al., 2010). At the same time,
Nicholas Economides and Joacim Tåg (2009) analyse stylized models of two-sided markets in an attempt to assess the welfare effects of net neutrality, and conclude that consumers are unambiguously worse off under net neutrality, while the effect on platform operators is ambiguous. Also, Florian Schuett (2010) provides an interesting survey of the economic literature related to net neutrality, which focuses on the incentive of market players to engage in conduct that allegedly would compromise the viability of the Internet. An analysis of the main economic features underlying the net neutrality debate is also available in Renda (2008, 2010).

From the standpoint of economists, regulating ex ante to flatten ISP practices would deprive society of an array of potentially welfare-enhancing transactions, in which players would reach different agreements based on their specific needs, and would meet at different levels of quality of service on the Internet. As clarified by Tim Berners-Lee a few years ago, “Net Neutrality is NOT saying that one shouldn’t pay more money for high quality of service. We always have, and we always will.” Brito et al. (2010) concur with this view – expressed also in Renda (2008) – when they state that “the practices that would be banned under the NPRM are likely, in most circumstances, to be welfare-enhancing. While it is possible to construct theoretical models in which economic welfare might be harmed, there is virtually no empirical evidence that such harm has occurred or is likely to occur in the future. Thus, it is extremely likely that the regulations proposed in the NRPM would harm consumers and competition and reduce economic welfare” (Schuett, 2010).

In greater detail, the economic literature is almost unanimous in considering second-degree price discrimination – including the sale of a fast lane on the Internet – as welfare-enhancing under reasonable assumptions. Papers such as Lee and Wu (2009) and Krämer and Wiewiorra (2009) provide useful guidance in this respect. At the same time, the ISP’s incentive to degrade the quality of competing products is more controversial: but this is exactly what competition laws are there for. Renda (2008) and Chirico et al. (2007) argue that under most circumstances antitrust law is already well equipped to tackle the problems that may emerge in this and other markets. Regulating ex ante to fix a problem that is common to several markets does not seem any different from throwing the baby out with the bath water. Again, the collective submission of several economists to the FCC confirms this view (Brito et al., 2010).

To be sure, the debate over the need to keep the telecom pipes as dumb as possible is not over. Many commentators, however, are starting to realize that the debate should either be stopped, or enlarged to all those players that play the role of platform operators, hold a share of the users’ attention and, consequently, can affect users’ decisions and alter competition in all complementary markets.
4.3 Unbundling and Network Neutrality in a Layered ICT World

As observed in the previous sections, the layered nature of new all-IP platforms adds several degrees of complexity to the already delicate assessment that needs to be made by the regulator wishing to impose an access obligation on a dominant network operator. One important consequence of the fact that next generation networks, for the most part, still need to be deployed in most countries around the world is the need to take into account the business case for investment, before the positive externalities associated with the availability of bandwidth can eventually be unleashed. The fact that broadband infrastructure conveys positive externalities to the whole economy, and to applications and content providers in the first place, must be adequately considered before a sustainable regulatory approach can be identified. The consequence of keeping a “copper era” regulatory approach in times of all-IP networks are at least twofold.

On the one hand, if the focus of the regulatory effort to boost competition in NGNs remains exclusively on the infrastructure layer, then the risk of undermining incentives to invest would be tangible. In particular, if policies at the infrastructure layer are not coordinated with those that affect the higher layers, the result might simply be no investment in infrastructure at all. This would be the case if Internet service providers knew that, following a massive investment in NGNs, they would be forced to keep their pipes “dumb”, as would be the case under a mandatory net neutrality scenario. The explanation is simple. On the one hand, under current regulatory arrangements and cost-based pricing, they would face very limited returns on their investment at the infrastructure layer. Moreover, any attempt to monetize the investment would be frustrated by the existence of intra-platform competitors that charge competitive retail prices, being able to rely on a regulated wholesale access charge. On the other hand, there would be no guarantee of any revenue coming from higher layers. Any attempt to charge for services would result immediately in lost customers, since “best effort” traffic would work for every player in the same way, and any price difference would result in switching. Finally, as an additional remark, mandatory net neutrality would further undermine incentives to invest since demand would likely be low for networks that are still prey to best effort traffic only.

In a different scenario, if the regulator approaches similar competitive problems in the same way on all the layers of the all-IP platform, there may be a significant risk of Internet regulation. Since bottlenecks and market power can emerge at all layers of the value chain, we should get ready for heavy regulatory intrusion into key service layers such as search. This is, to some extent, already happening with antitrust probes into online search firms and online advertising companies, but may be exacerbated to reach instances of functional
separation of multi-product giants that act as platform operators. And with the emergence of cloud computing, there may be scope for preventing exclusionary abuses on the cloud by granting open access to all products. Under such a scenario, it would not be strange to see calls for mandating open access to platforms such as Apple’s App Store or Google Android. The absence of product differentiation and innovative business models that this situation would create makes this scenario very undesirable from a social welfare perspective.

As a result, the only solution for a regulator is to fine-tune policy actions with a focus on balancing the incentives of all players involved. The best regulatory approach for the infrastructure layer, thus, may not always be unbundling, and certainly will not be unbundling whenever mandatory network neutrality is pushed to the extreme (with the sole exception of publicly funded networks). This circumstance makes the work of sectoral regulators even more difficult, and the case for unbundling as the mainstream model for future telecoms regulation even more unlikely.

BIBLIOGRAPHY


372 Regulation and economics


Further Readings


Gerrit De Geest - 10.4337/9781782540465.00016 Downloaded from Elgar Online at 11/26/2018 03:53:36PM via free access