International Telecommunication Union

GSR 2013 Discussion Paper

Interconnection charging models in a national broadband network environment

Work in progress, for discussion purposes

Comments are welcome!

Please send your comments on this paper at: gsr@itu.int by 19 July 2013.

Contents

1 INTERCONNECTION CHARGING MODELS IN A NATIONAL BROADBAND NETWORK ENVIRONMENT

David Rogerson: Director, Incyte Consulting

Introduction

In the early 2000s, as the so-called "Dotcom Bubble" burst, incumbent telecommunications operators sought to take advantage through the development of Next Generation Networks (NGNs)[1]. These are packet-switched networks based on the Internet Protocol (IP) which maintain the best features of legacy circuit-switched voice networks in terms of guaranteed quality of service, while improving efficiency, lowering costs and facilitating the introduction of new services. NGNs were the established telecoms players' response to the rising tide of the Internet, designed to protect their market position and their profitability for as long as possible.

The success of the Internet derived from open unrestricted access to services created by users and located at the edge of the network. NGNs were the telecom companies' attempt to replicate the architecture of the Internet while building on the very high reliability and predictable quality of service provided to end-users over their circuit-switched networks. This counter-strategy depended on establishing effective interconnection between all the NGNs so as to replicate the open-access regime of the Internet, while allowing the telecom operators to maintain control of the infrastructure and continue with the usage-based charging system from which they derived their profitability. For the purposes of this paper, it is sufficient to note that the main aspects NGN and the Internet hold in common are the use of IP transmission technology and the decoupling of service provision from network facilities; and the main characteristic that separates them is the promise of end-to-end quality of service in the NGN and the absence of any such promise in the "best efforts" Internet.

At this critical juncture, just over 10 years ago, IP interconnection was compared with the laws of thermodynamics[2]. Just as the laws of thermodynamics describe fundamental truths of the universe in terms of basic elements such as heat, energy and matter, so the parallel "laws" of IP interconnection talk of the basic elements of revenue, usage and market power that affect the telecommunications universe[3]. Both systems – the universe and telecommunications markets – are complex but not entirely chaotic, and it was postulated that the rigorously defined laws of

[1] Both terms, "Internet" and "NGN", are open to multiple interpretations. In particular, as described in Section 1.2, the term NGN has undergone significant development since it was first used in the 1990s. The ITU's definition of NGN can be found at: http://www.itu.int/dms_pub/itu-t/oth/1D/0B/T1D0B0000010003PDFE.pdf.

[2] Rogerson D, Horrocks J, Hin J, Lavender T: "IP Interconnect: Commercial, Technical and Regulatory Dynamics"; Ovum (2002)

[3] The relationship with the Laws of Thermodynamics is one of metaphor, and is used to throw new light on the subject and to help understanding. However, it should be remembered that all metaphors break down if pushed too hard, and this one is no exception.

thermodynamics may have parallels in the telecom world that would help guide market participants through the turbulent phase of transition from circuit-switched to IP interconnection. The postulated Laws of IP Interconnection are shown in Table 1.

The objective of this paper is to examine the extent to which these "laws" of IP interconnection have stood the test of time. Each of them purported to tell of an immutable truth that should shape the thinking of an industry in a time of transition: have they proved to be accurate over the intervening years of rapid change and further technological progress? Have these principles stood the test of time, and what have been the consequences? This paper will seek to answer such questions and use the prism of the "laws" to throw fresh light on the challenges of interconnection in a national broadband network environment.

Table 1: The parallel laws of thermodynamics and IP interconnection

Law number	Law of Thermodynamics	"Law" of IP Interconnection
0	If object A is in thermal equilibrium with B, and B is in thermal equilibrium with C, then object A is in thermal equilibrium with C	If network A is interconnected with B, and B is interconnected with C, then network A is in effect interconnected with C
1	The change in internal energy of a system is equal to the amount of heat added to the system less the work done by it. This means that energy cannot be created, it can only be transferred from one system to another.	Within any interconnected system, the total profitability is related to the amount of usage of the system and the relevant tariff. This means that profitability cannot be increased without an increase in retail price or usage, no matter how fast IP is installed.
2	Within any closed system, the entropy (that is a measure of the disorder or random distribution of matter throughout the system) always increases with time.	Within any IP system, the complexity (that is a measure of the disorder or random distribution of packets and usage throughout the system) always increases with time.
3	The universe will gradually move towards a state called absolute zero, at which time all energy and matter will be randomly distributed across space.	The IP world will gradually move towards a state in which market power is removed, at which time usage and profitability will be randomly distributed throughout the system.

Source: Ovum

1.1 The Zeroth Law: interconnection is transitive

After the first three laws of thermodynamics were established it was decidedly inconvenient when another, more basic, law was proposed. Showing the creativity of engineers it was coined the Zeroth Law so as to leave the other three undisturbed. In a similar manner the Zeroth Law of IP Interconnection is easily overlooked as it seems obvious: there is no need for all networks to be directly interconnected so long as there is sufficient interconnection that any-to-any connectivity can be assured.

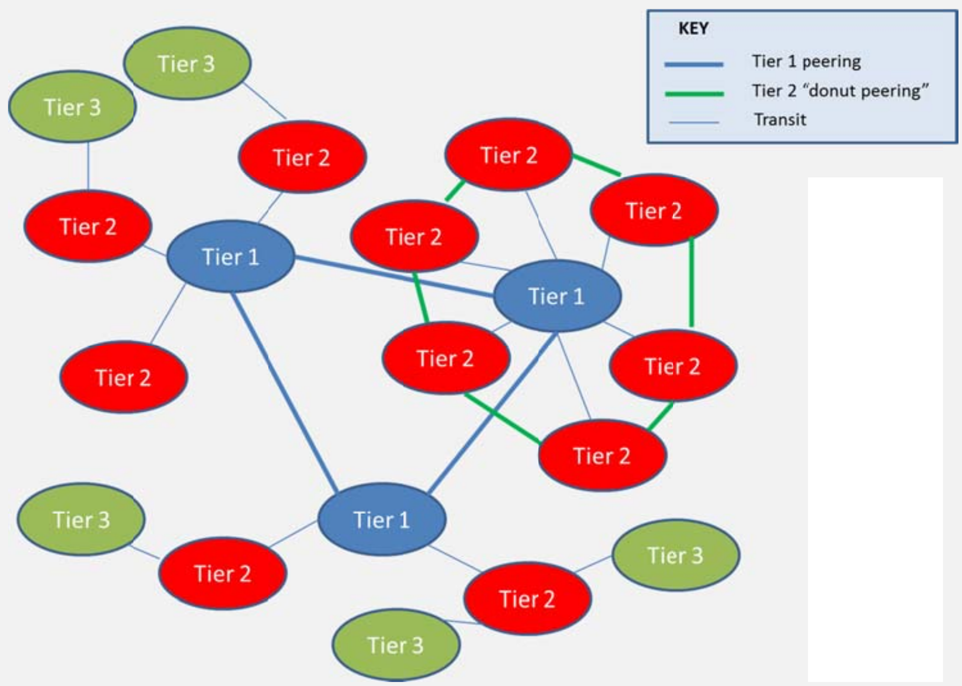
1.1.1 Peering and Transit

In the circuit-switched world of telecommunications any-to-any connectivity has been imposed as a regulatory requirement and terms of interconnection have been carefully controlled to ensure that network effects are enjoyed at the total system level and are therefore delivered to users in terms of greater amenity and reach. Any-to-any connectivity ensures that network benefits are not privatised and hence also limits the exploitation of market power. None of this has happened in the public Internet, and yet full connectivity has been achieved through two forms of interconnection[4]:

- **Peering:** in which two interconnecting systems are treated as (approximately) equal, with the result that no billing is required between them. In practice the networks will not be exactly equal, but variations in matters such as the volume and direction of traffic and the geographical reach of the network, will be held within acceptable bounds. Allowable variations may be determined in advance and specified in the peering contract.
- **Transit:** in which there is a considerable disparity in the size or value in the interconnection relationship, with the result that the larger network treats the smaller one as a customer, charging it for handling both incoming and outgoing traffic. Transit is the default option, providing connectivity to the full Internet for those operators which do not have the scale or scope to enter into peering relationships.

These two forms of interconnection are shown in Figure 1. Together they ensure full interconnectivity as a result of the transitive properties described in the Zeroth Law: the larger transit networks perform the function of linking smaller networks that could not commercially justify direct interconnection. The system is thus highly effective and efficient, achieving a full mesh with a minimal number of separate interconnection agreements.

Figure 1: Peering and transit interconnection arrangements[5]



Source: Author

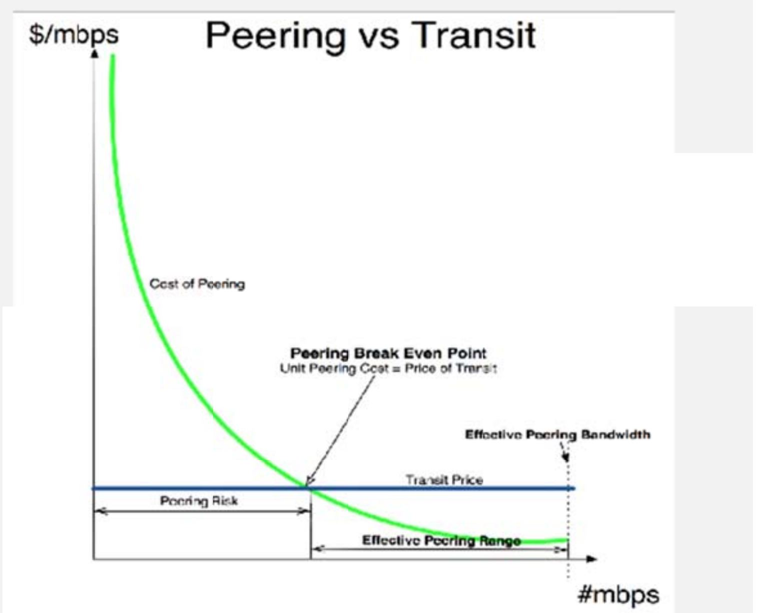
[4] The interconnection arrangements of the Internet that are summarised here were fully explored at GSR2012 in the paper by Dennis Weller entitled "Blurring Boundaries: Global and Regional IP Interconnection".

[5] Each "tier" in this diagram refers to a group of ISPs that have peering relationships. The largest ISPs are known as Tier 1, and the Tier 2 ISPs are interconnected to them via a transit agreement.

In the early days of the Internet, the system of peering and transit was fiercely contested, particularly by those operating within the telecommunications paradigm. It was considered unfair that peripheral networks were forced to pay to belong to the system; in practice this meant funds flowing from developing countries, through major centres in Europe and Asia, ultimately to reach the core of the Internet in the USA. Such an approach appeared to reverse the subsidies that were embedded in the International Accounting Rate system of international telecommunications at that time.

While there was some truth in these claims, with hindsight they were over-stated, as more and more companies, not just the Tier 1 ISPs[6], have been able to establish peering relationships. Peering, even where it is settlement-free, has a cost in terms of transmission to the peering point, co-location, ports and equipment. There is also the operational cost of building and maintaining the peering relationship. In effect transit interconnection involves paying the transit operator to conduct the peering relationship on behalf of its customers, since those customers themselves are too small or remote for peering to be cost-effective. Over time, as the volumes of traffic grow, peering relationships begin to be cost-effective as shown in Figure 2.

Figure 2: Cost effectiveness of peering and transit



Source: Dr Peering

Even if in a static sense the criticisms of transit interconnection were valid, the dynamism of the Internet has ensured that any unfairness was short-lived as work-around solutions have been established. Peering has rapidly replaced transit as the norm with the result that the vast majority of traffic no longer touches the major global backbone networks. 99.5% of international IP interconnection (transit or peering) agreements are reached "on a handshake" without written

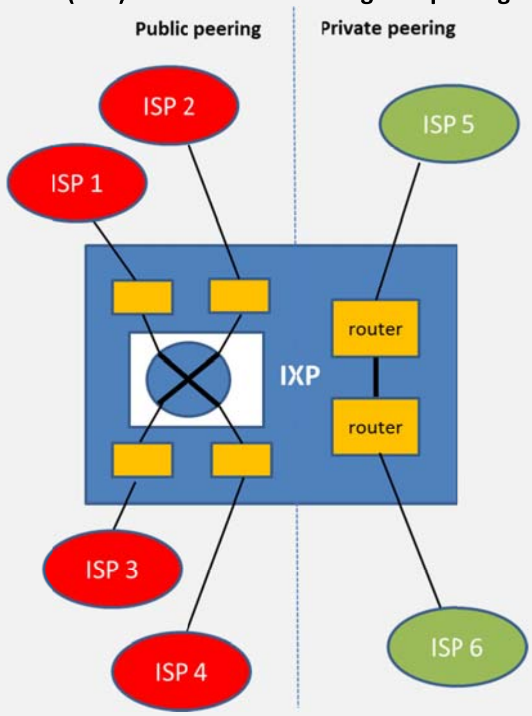
[6] A Tier1 ISP is an ISP which has access to an entire Internet Region solely via free and reciprocal peering agreements.

contract and only 0.27% of agreements were asymmetric (e.g. included a payment, or imposed minimum quality terms)[7].

This transition has been achieved by the confluence of a number of factors, the motivation for each being to reduce the overall cost of interconnection arrangements as Internet traffic volumes have grown:

- Growth of regional peering. In the early days of the Internet peering was restricted to the largest ("Tier 1") backbone providers, but there is now a much greater range of peering options among lower-level regional providers (sometimes called "donut peering").
- Growth and distribution of Internet Exchange Points (IXPs). IXPs are locations where multiple Internet service providers (ISPs) meet to effect peering, as shown in Figure 3. Although each ISP has to bear its own costs of transmission to the IXP, these costs are spread across multiple peering relationships achieved at the same location, thus improving the affordability of peering.

Figure 3: Internet Exchange Points (IXPs) allow cost-effective regional peering



Source: Author

- The development of regional content and content distribution networks (CDNs). The development and caching of content within regions, including developing countries, creates additional demand (more desirable content is more easily accessible), reduces costs (as transit interconnection is no longer required) and improves service quality (latency is reduced as there are fewer "hops" to reach cached content). It is also a key component of

[7] Weller, Dennis, "Blurring Boundaries: Global and Regional IP Interconnection"; ITU GSR2012

ensuring that IXPs have sufficient bargaining power with bandwidth providers so as to achieve affordable prices, which is especially important in developing countries.

- IXPs also act as a spur for local investment in content and related industry, in much the same way that airport hubs generate significant economic activity within their hinterland. In East Africa, for example, a liberalised technology sector has enabled Nairobi's development as a regional IXP and has resulted in much digital content being cached in that country; while in neighbouring Ethiopia, the government has invested heavily in the development of Addis-Ababa as a regional airport hub[8].

1.1.2 Any-to-any connectivity requirements

Experience since the start of the millennium has tended to corroborate the Zeroth Law with respect to the public Internet. There is clearly no need to impose any-to-any interconnection requirements between these networks (or the applications that run over them) as commercial demands are sufficient to achieve full interconnectivity. But where does that leave the case for requiring any-to-any connectivity for voice services, be they provided over the traditional PSTN or an NGN[9]; and should such requirements be extended to any other services provided over NGNs?

There is a public interest case, on the grounds of both economic efficiency and social inclusion, for requiring interconnection between circuit switched networks for any-to-any connectivity so as to ensure the delivery of end-to-end voice services. Such a requirement is commonplace in operators' licences or in telecommunications legislation throughout the world, and is based on the economic argument that welfare is maximised when subscribers on one network can call (and receive calls from) subscribers on all other networks. For basic voice services, the any-to-any obligation is imposed for the avoidance of doubt, in particular because larger networks may have an incentive not to provide it for anti-competitive reasons. However, the requirement is often imposed on all operators and not just those with Significant Market Power, so that all users on all networks are protected. There is no reason why these requirements should be withdrawn when voice services are carried on NGNs rather than circuit switched networks: the change of network technology does not alter the case for any-to-any connectivity.

As and when consumers come to rely on over-the-top voice applications on the Internet, such as Skype, rather than the traditional voice services of the PSTN with their defined quality of service standards, the need for continuing regulation will be reduced and may eventually disappear. However, these circumstances are still some way off. Ovum, for example, has suggested that "reports of the death of the [PSTN] telephone have been greatly exaggerated" and VoIP will actually replace only 6.9% of telecom operator revenues between 2012 and 2020[10]. In such circumstances it is unsurprising that no country has yet taken the step fully to deregulate voice communications.

---

[8] The contrasting approaches to economic development in Kenya and Ethiopia are discussed in: http://www.economist.com/news/special-report/21572379-ideological-competition-between-two-diametrically-opposed-economic-models-doing-it-my

[9] The case of VoIP is dealt with later as it primarily concerns quality of service. As far as any-to-any connectivity is concerned, VoIP relies upon the full connectivity of the public Internet, as described in the previous section.

[10] Green J. and Obiodu E. "The Future of Voice", Ovum, 2012

Before extending any-to-any connectivity requirements services other than voice the following factors need to be taken into account[11]:

- the extent of competition in the supply of the service. If there is no dominant supplier then there is little danger of anti-competitive behaviour (avoiding interconnection so as to gain competitive advantage) and every chance that the market will reach any-to-any connectivity outcomes without regulatory pressure, as it has with the public Internet.
- the extent of network externalities. How much would end-users benefit from being able to connect to all other end-users and not just to those on their own network? How much would end-users benefit from being able to access any service/server and not just those on their own network? Only if such externalities are high would a regulatory requirement for any-to-any be justified.
- the costs of the any-to-any requirement. These include the direct costs associated with establishing gateways and maintaining interoperability, especially in the case where standards are yet to be fully established, and hence the danger of stranded investment is high. Account should also be taken of the opportunity costs – e.g. in the absence of any-to-any requirements, competition between network operators may be increased as they use access to innovative applications and/or connectivity with other network subscribers as a means of attracting new customers.

It is not clear that any service, other than voice, would pass these tests; it is even doubtful that voice would pass the tests in markets with highly developed over-the-top voice applications. SMS, MMS are possibilities, but most regulators have declined to regulate these services. Real time video is another possibility, but the scale of the externality effect is likely to be low compared with voice, and unlikely to justify the significantly higher baseline network costs required for such a service. Few other applications are provided on an end-user to end-user basis, with the result that any-to-any connectivity would bring little public benefit. Furthermore, absent the any-to-any requirement, there may be enhanced competition and innovation in applications as each operator uses these services to attract new subscribers to its network.

1.1.3 Regulating how interconnection takes place

Not content with the principle of any-to-any connectivity, some jurisdictions have established specific arrangements for how interconnection takes place. For example, in some countries, all interconnected traffic must pass through Interconnect Exchange (ICX) operators, which are licensed separately for this purpose. These arrangements may have been established for good reasons (e.g. to crack down on the means of avoiding payment of usage-related licence fees), but their effect may be to fossilise one interconnection architecture, as the ICX operators have a rent-seeking vested interest in maintaining the status quo. The risk is that this prevents progress towards more efficient arrangements over time and the public interest and consumer welfare are injured in the process.

Another arrangement that is regressive is insistence that traffic be converted to time division multiplexing (TDM) before being passed between networks. In a predominantly circuit-switched

[11] The argument presented here on the applicability of any-to-any connectivity to services other than voice is derived from the report *Next Generation Networks: Next Generation Regulation?* prepared by Plum Consulting for OFTA in Hong Kong. It is available at: http://tel_archives.ofca.gov.hk/en/report-paper-guide/report/rp20120305.pdf

world this had some virtue, as the majority of traffic used TDM, so the costs of conversion from TDM to IP were greater than the reverse. But as IP becomes the transmission protocol of choice, these economic arguments are increasingly being turned on their head, and the only reason for staying with TDM-based interconnection is to protect the interests of incumbent operators which would otherwise have to bear the costs of converting legacy circuit-switched traffic to IP. Regulators have been lacking in leadership here: some have investigated the issue, but they have so far refrained from taking decisive action by withdrawing requirements for IP-to-TDM conversion at the point of interconnection.

1.1.4 Conclusions

There is a reasonable presumption that any-to-any connectivity will be achieved and maintained throughout the public Internet. Such conditions could potentially apply to all networks including the PSTN and NGNs, but to provide absolute certainty it is reasonable to maintain the regulatory requirement for any-to-any interconnection for voice services within the PSTN/NGN environment. However, these regulations should not be extended to other services unless strict public interest tests are met. Over time it is more likely that the any-to-any requirement could be withdrawn rather than extended to other services or applications. Apart from the safeguard of any-to-any connectivity, there should be no need to restrict who can provide interconnection or to regulate interconnection architectures.

1.2 The First Law: profitability increase depends on growth

1.2.1 The development of NGNs

The original NGN concept was of a managed IP network and comprised:

- A closed network with vertical integration between transport and services
- Interconnection via gateways which carry out security, charging and signalling functions
- Services delivered to customers by each NGN operator and by third parties using a controlled Application Programming Interface (API)
- Large-scale emulation of traditional services and a range of new services
- A layered approach with access, transport, service control and application layers – allowing integration of fixed and mobile networks so as to enable a common transport network and common communications services.

This concept is shown in Table 2 in contrast to the concepts underpinning the Internet. It was hoped that profitability could be increased by a combination of cutting costs through the concentration of all services on a single network platform while simultaneously countering competition from the Internet by providing improved service quality.

Table 2: NGNs versus the public Internet

Factor	NGN	Public Internet
IP-based?	Yes	Yes
Interconnection	Closed via gateways for security, quality of service and charging purposes.	Open
Intelligence	Centralised and controlled by the NGN operator.	At the network edge and implemented by end-users.
Innovation	By NGN operators and third parties over controlled APIs.	By users at the network edge – innovation without permission.
Ability to roll out services	Local	Global
Management of quality of service	Through central control of resource allocation for intra-NGN services. Competing solutions for inter-NGN services, with limited implementation to date.	Best efforts service. Build additional capacity to avoid congestion. Traffic management at network edge.

Source: Plum Consulting

Progress in implementing NGNs has been both limited and variable for two main reasons:

- The technical challenges inherent in replacing a range of traditional networks with a single NGN has been greater than originally anticipated with the result that projected cost savings have generally not been realised.
- The new services, which NGNs were designed to carry and which were supposed to help generate a return on the investment, have largely been implemented on the Internet instead. While some high-quality video applications are carried over NGNs, the Internet now generates over 99% of telecommunications traffic in the developed world (see Figure 4 below). For most applications in the consumer market, and increasingly for many business applications, the best effort standard is good enough. For example, most businesses rely on the Internet for communicating with suppliers, customers and intermediaries, such as financial institutions, and for these purposes generally accept the best efforts service. Higher quality services are the exception and are likely to remain so.

Despite these issues the implementation of NGNs has continued. The success of the Internet means that conversion to IP technology is the only available option. Increasingly few vendors support traditional circuit switched technologies such as public-switched telephone network (PSTN), frame relay and asynchronous transfer mode (ATM) networks, with the result that these networks are increasingly expensive to maintain. In that sense the cost savings that were predicted for NGN have come true: migration to NGNs has not cut costs in absolute terms, but it has prevented an ever-increasing maintenance bill on legacy networks.

A typical NGN example is provided by BT in the UK (see Box 1). This example shows how the concept of an NGN has changed over time. There is now much less talk of service migration, cost savings and managed IP services: the NGN is now primarily about providing fibre backhaul to support the growth of the Internet and, through related Next Generation Access (NGA) projects, providing higher access speeds to retail customers.

Box 1: BT's 21st Century Network

In 2005 BT announced plans to roll out its 21st Century Network (21CN) across the UK. These plans were designed to deliver a complete replacement for up to 17 networks by 2011 and save £1 billion per year in operating costs. In practice this vision proved overly ambitious and many legacy services were never transferred to the new network, although in some cases roughly equivalent services were launched on the NGN. In recent years BT has stopped referring to the 21CN, and the term being replaced by Next Generation Broadband (NGB). This is a £2.5 billion fibre-to-the-cabinet investment programme that aims to pass two-thirds of UK premises by 2014, and may extend to 90% of premises on the back of Public Private Partnership arrangements in remote and rural areas. In April 2013 BT announced that its NGB network had passed 15m million premises – more than more half of all premises in the UK.

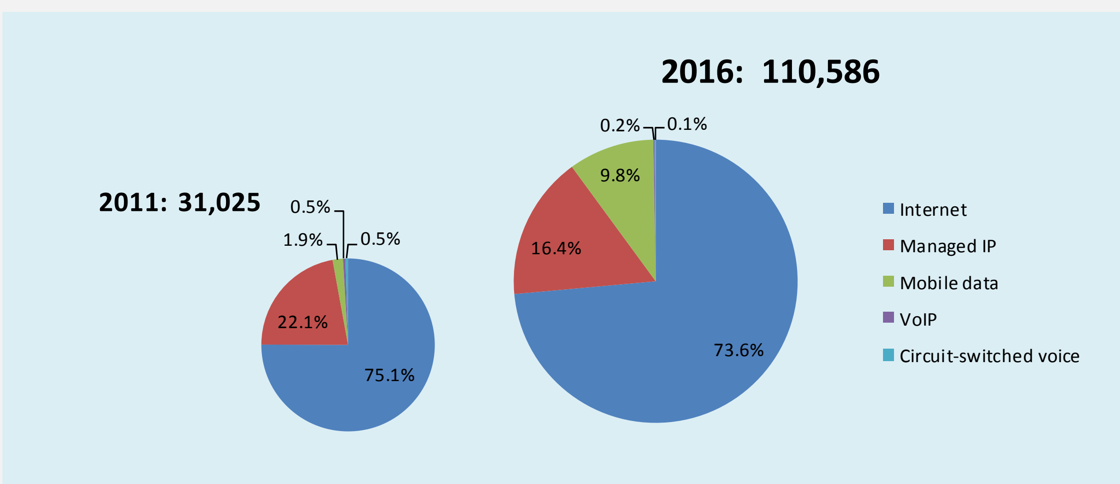
Source: Author and BT (http://www.btplc.com/ngb/)

1.2.2 Why NGNs have struggled

One of the perceived advantages of the NGN was its ability to enable telecom operators to deliver a wide range of new, revenue-generating services from within a "walled garden" created by gateways to other networks. But, in practice, the Internet has delivered the vast majority of these new services which have substituted for the revenues anticipated by the telecoms operators. The Internet has achieved this because, unlike NGNs, it locates intelligence at the network edge in the servers and clients of end-users, and provides an open transport network to link them together, allowing users to innovate without permission and generate a far richer array of applications than the telecom operators ever could.

The net result is that the Internet has grown far more strongly than NGNs over the past 10 years and there is no sign of that situation changing in the future. Figure 4 illustrates how the public Internet has come to dominate telecommunications. It compares the traffic carried over the Internet with managed IP traffic and circuit-switched traffic for the period 2011-2016, based on recent Cisco forecasts[12].

Figure 4: Global traffic forecasts in PB per month



Source: Author, using Cisco forecasts (with additional data from Google)

[12] See: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

It is possible that in the future a managed layer within the public Internet may emerge which may further weaken the business case for NGNs. But this seems unlikely given the global nature of the public Internet and the consistent improvements in "best efforts" performance that it has delivered. A more likely development – and the one which underpins Cisco's forecasts – is that NGNs will continue to deliver services which require guaranteed quality of service, principally business data and high-quality video, but these will be a diminishing proportion of total services, while the public Internet will continue to deliver the majority of applications on a best efforts basis.

1.2.3 Implications for interconnection

Although NGNs were designed to supply managed IP services for the 21st century, in practice they occupy a small corner within the market for IP-based telecommunications services as a whole. This corner happens, for reasons related to legacy circuit-switched networks and PSTN voice services, to be the part of the market most prone to regulated interconnection. However, as the dominance of the public Internet continues to grow, there is the likelihood of unforeseen and potentially damaging consequences if ex-ante regulation is simply transferred to the replacement NGNs.

This situation has arisen largely because NGNs have fallen foul of the First Law: they have failed to deliver the promised growth in service revenues, without which there is no means of increasing long-term profitability. Even if NGNs have reduced costs through platform integration (itself a disputable point) such cost reductions have failed to stem the double-hit on revenues as price levels and traffic volumes have both fallen in response to "free" Internet applications.

Regulators need to be careful in their reaction to NGNs. At the service level, if anything, regulation ought to be peeled away as the presence of alternative Internet applications will on its own be able to control the market behaviour of NGN operators (i.e. managed IP service providers). The burden of proof therefore lies with the regulator to demonstrate why any interconnection obligations are required, and to ensure that the pace at which they are introduced is consistent with the pace of migration to NGNs. Equally, however, regulators need to ensure that the broadband networks that have been developed as NGNs, including both passive and active infrastructure, do not become a new source of monopoly rent. This infrastructure needs to be made available on open access principles to all service providers, be they ISPs or managed service providers.

1.3 The Second Law: complexity always increases over time

1.3.1 Evidence of increasing complexity in IP networks

The rapid growth and development of IP networking over the past 10 years is undeniable. It is also, according to the Second Law, inevitable. The change of scale and complexity in the broadband/Internet ecosystem has been staggering as just a few facts indicate.

- There has been significant and sustained growth in global IP traffic. Table 2 provides Cisco's forecasts which show that there will continue to be strong global growth at least until 2016, but the rate of increase is declining slightly, to around 29% per annum. The figures also demonstrate the continued strength of the fixed Internet although the rate of growth is much higher for mobile data. Lower growth is forecast for managed IP networks.

Table 2: Global IP traffic developments

	2011	2012	2013	2014	2015	2016	CAGR
By Type (PB[13] per Month)							
Fixed Internet	23,288	32,990	40,587	50,888	64,349	81,347	28%
Managed IP	6,849	9,199	11,846	13,925	16,085	18,131	21%
Mobile data	597	1,252	2,379	4,215	6,896	10,804	78%
By Segment (PB per Month)							
Consumer	25,792	37,244	47,198	59,652	76,103	97,152	30%
Business	4,942	7,613	9,375	11,227	13,130	7,613	22%
By Geography (PB per Month)							
North America	10,343	14,580	17,283	19,796	23,219	27,486	22%
Western Europe	7,287	10,257	13,026	16,410	20,176	24,400	27%
Asia Pacific	10,513	14,792	18,976	24,713	31,990	41,105	31%
Latin America	1,045	1,570	2,333	3,495	5,208	7,591	49%
Central and Eastern Europe	1,162	1,673	2,290	3,196	4,419	5,987	39%
Middle East and Africa	384	601	903	1,417	2,320	3,714	57%
Total (PB per Month)							
Total IP traffic	30,734	43,441	54,812	69,028	87,331	110,282	29%

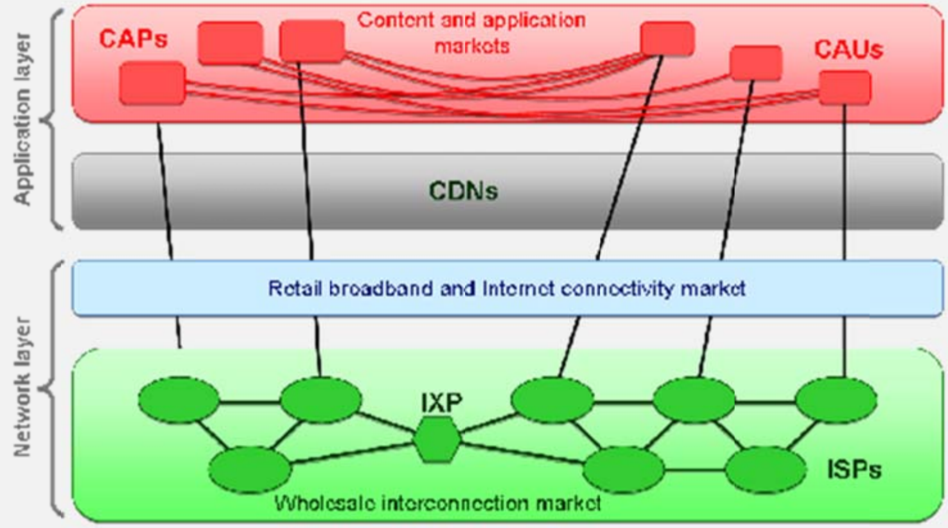
Source: Cisco (for weblink see Footnote 9)

- The number of players in the broadband value chain has increased. Figure 5 describes the principal categories of players in the market, but individual companies may perform more than one function. Perhaps the most significant change in recent years has been the emergence of Content Distribution Networks (CDNs). These are service providers designed to enhance quality of delivery of Internet content, which they achieve through local caching and greatly increased amounts of direct routing. CDNs generate revenues from Content and Application Providers (CAPs) and they generally price their services per Mbit, but they also offer value added services such as digital rights management[14] and regional restricted content.

[13] A Petabyte (PB) is equal to 10^{15} bytes

[14] Digital rights management is a range of technologies used to control access to and usage of digital content.

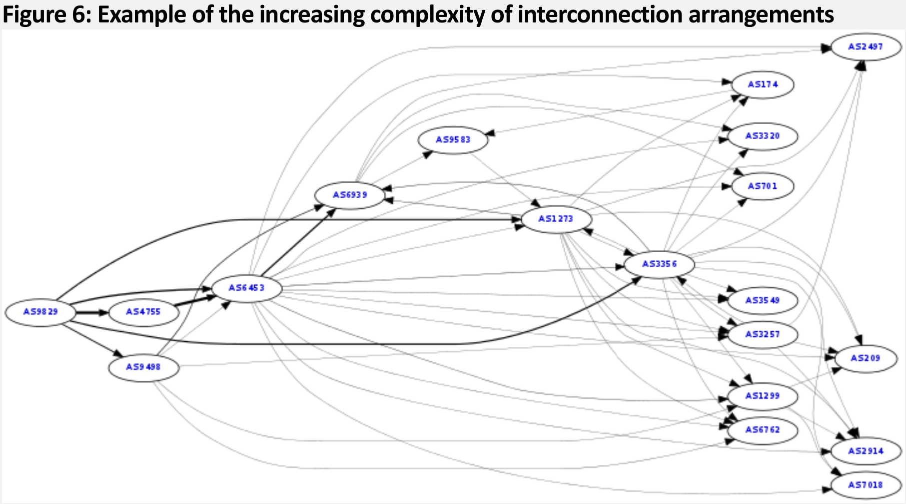
Figure 5: Increasing complexity of the broadband value chain



Source: BEREC

- The move away from transit interconnection, significantly achieved through the growth of CDNs, has increased the variety and complexity of interconnection arrangements. Figure 6 provides one example based on transit (dark lines) and peering (light lines) connectivity from one node in the BSNL network in India[15].

Figure 6: Example of the increasing complexity of interconnection arrangements



Source: Hurricane Electric (as of 23 April 2013)

- The variety of services and applications carried over broadband networks has mushroomed. For example, the number of applications in the Apple AppStore has risen from 50,000 in June 2009 to 775,000 in January 2013[16]. More generally, cloud services have taken off with

[15] See: http://bgp.he.net/AS9829

[16] See: http://ipod.about.com/od/iphonesoftwareterms/qt/apps-in-app-store.htm

56% of Internet users having webmail accounts, 34% storing photos online, 29% using online applications, and 5% backing up their hard disks online[17].

- Traffic patterns have changed as new applications have taken off. Video streaming is by far the most bandwidth-intensive application, accounting for more than 50% of traffic volumes; this along with file sharing means that the majority of traffic is no longer person-to-person, and the peak hour loading of the network has increased significantly, representing as much as 50% of total traffic. These trends, however, are asymmetrical: they affect downloads rather than uploads, and person-to-person communications continues to dominate upstream communication.
- Unit costs have tumbled both as a result of volume growth and technological development e.g. rapid reductions in the costs of routers and dense wave-division multiplex (DWDM) equipment. The cost of mobile data has been falling particularly rapidly as spectrum efficiency improvements have lowered costs per MB.

1.3.2 Murphy's Law and the need for simplicity

The Second Law of thermodynamics has often been paraphrased in what has become known as Murphy's Law: "that which can go wrong will go wrong". In systems with ever increasing scale and complexity, there is an imperative not to add unnecessarily to the complexity as this is only likely to increase the chances of things going wrong. Undue intervention in the market is a temptation of which regulators need always to be mindful, but especially so when there is a high degree of flux in the market. This situation clearly applies to Internet and NGN interconnection, and may be demonstrated in relation to the specific example of quality of service regulation.

1.3.3 Regulating for Quality of Service

Quality of service is almost always a trade-off against cost of service. Consumers want to maximise quality and to minimise cost. Suppliers cannot achieve both of these goals simultaneously, but in a perfectly competitive market they will find the right balance between cost and quality that maximises consumer welfare.

If a market is not effectively competitive – in other words where there is a supplier with significant market power (SMP) – there may be a need to establish minimum quality of service standards, in just the same way as there may be a need to establish maximum prices. In the legacy circuit-switched world this was indeed the case, and minimum quality standards for voice services were established at an international level, by the ITU and others.

As networks transition to NGN, these quality standards continue to provide a baseline for voice services. In many countries the voice market is sufficiently competitive that any operator which fails to provide the baseline level of quality demanded by consumers will tend to lose custom to other operators that do meet this quality standard. Operators that fail at least to provide equivalent quality standards to their existing services using legacy networks will find it hard to persuade customers to switch to IP-based voice services. Customers might accept *best efforts*, but that does not mean *any efforts* are good enough. They have clear expectations from past circuit-switched experience. However, in developing countries users generally have not had a high-quality

[17] See: http://www.cloudhypermarket.com/whatiscloud/CloudUptake

experience of fixed telephony, with the result that they are all the more receptive to IP-based services.

This suggests that there is no need to establish minimum quality standards for voice services provided over the NGN. There are also good practical reasons for regulatory forbearance:

- It will be difficult to set the economically optimum minimum quality standard through regulation. While theoretically there is a cross-over point in the cost-quality dynamic at which consumer benefit will be maximised, this point is impossible to define a priori. There is certainly no reason to believe that the legacy standards from circuit-switched networks are set at the right level, and especially not at the right level for an IP world. Consumers may well be prepared to trade-off quality against further price reductions[18], or, in some cases, they may be willing to pay more for higher quality of service. The cross-over point is best determined by the competitive market.
- Quality of service standards on NGNs are not yet sufficiently developed to enable regulators to define and impose a minimum quality standard. Different standards bodies (e.g. ITU Study Group 12, IETF, GSMA, i3 Forum, ETSI 3GPP) are working to define classes of service quality on NGNs, but as yet these standards are still under study. It is clearly too early to set any kind of minimum quality standard, even for voice services.
- The existing quality standards for IP provide traffic management at the network layer but do little to describe or determine the full user experience. This makes the standards appropriate for deployment within IP networks, where they have been successfully implemented for maybe 10 years, but it also makes them less attractive for use at the application layer (e.g. for voice services). The high transaction costs of negotiating and monitoring quality across the point of interconnection also militate against end-to-end quality measures, as does the continuing improvement of the "best efforts" Internet to meet evolving customer demand.

An alternative regulatory approach would be to:

- Retain minimum end-to-end quality of service standards for circuit-switched voice. These will act as a baseline for operators transitioning voice services to IP networks.
- Encourage industry to establish guidelines for minimum quality of service standards for voice communications on NGNs, and to make proposals to the regulator. This work should be carried collaboratively (e.g. though an industry forum) and should wherever possible build on international standards, such as Internetwork Packet Exchange (IPX).
- Monitor what happens, in particular conducting surveys to gauge consumer reaction to service quality, and intervene in the future if necessary and where feasible.

All the difficulties that apply to setting quality standards for voice services apply equally to other services that may be provided over NGNs. There is a further difficulty: whereas voice services are clearly understood and have a legacy quality of service standard associated with them, this is not the

[18] The success of best efforts VoIP services such as Skype appears to support this point.

case for other applications. This makes it very difficult to apply regulation in this area, although regulators should maintain a watching brief on industry developments.

Most of the international standards bodies, including ITU, are working with the Internet Engineering Task Force (IETF)'s DiffServ classification of services (conversational, streaming, interactive, background), so that quality standards may be applied across a group of relatively homogenous applications. As ITU SG12 has stated[19], it is crucial to be able to measure new parameters such as packet loss and jitter, and know their user impact. To this end IPX has defined a range of quality parameters (e.g. availability, jitter, packet loss, delay) characterising each service classification, thus providing a proto-type for bringing QoS-based interconnection to the market, based on principles of open standards and flexible charging mechanisms.

However, international standards activity regarding costs and charges for guaranteed quality of service has a long way to go. A recent European Commission report[20] concludes that:

> *The mapping of IP traffic in different classes of quality is an important first step for enabling the differentiation and management of quality also between IP networks. However, this feature alone is not able to guarantee an absolute end-to-end quality (from an end-user to another end-user, or from a service centre to an end-user). In fact, the interconnection is not aware of the service (e.g. SIP sessions) transported by IP packets, so it's not possible to give any guarantee on the level of quality associated with each single service. It is reasonable to say that the diversification of IP traffic at transport level in classes of quality enables a relative quality of IP services/applications, which means that the quality could be associated to an IP traffic flow shared by many applications and services requiring the same quality objectives at IP transport level.*

It is difficult to conceive of any circumstances in which, or any services for which, a minimum quality of service standard would need to be imposed via regulation. The concern that operators with significant market power (SMP) may gain competitive advantage through externality effects is best dealt with through a non-discrimination requirement. This would mean that wholesale services need to be offered to interconnecting parties in the same combinations of quality and price as to the service provider's own retail operations ("equivalence of inputs").

1.4 The Third Law: the gradual erosion of market power

1.4.1 Market power in circuit switched networks

Before telecommunications was disrupted by the Internet the received wisdom was either of a natural monopoly or (following the advent of competition in the 1990s) of an imperfectly competitive market in need of regulation. Current regulatory best practice is based on the determination of SMP or dominance and the implementation of specific ex-ante remedies to protect weaker competitors from any abuse of market power. These preventive measures, as distinct from standard ex-post Competition Law remedies, are justified because network industries offer

[19] For the latest news from SG12 see: http://www.itu.int/en/ITU-T/about/groups/Pages/sg12.aspx

[20] Working Group 2 report on the Technical Framework for Digital Delivery, 13 July 2011, p20-27 at http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=7211

substantial economies of scale and scope, and thus naturally lend themselves to a small number of market players.

The regulation of access and interconnection has been the main focus of attention. First, incumbent or SMP operators must have the obligation to provide interconnection to any other operator[21]. Absent such regulation, there is a high risk that the dominant provider will simply refuse to supply interconnection with the result that would-be competitors are unable to supply retail call services to the majority of subscribers. The loss of utility would be so great as to make the new entrant's business unviable.

Secondly, the terms of interconnection need to be regulated. The precise nature of the obligations will depend on the services being offered and the extent to which the market on its own can constrain the SMP operator's behaviour. Typical requirements include transparency, non-discrimination, equivalence of inputs and cost orientation. The aim is to ensure that the interconnection service offered to competitors is such that the latter may compete on equal terms in the downstream retail markets that depend on the wholesale inputs provided through interconnection. So, for example, call termination rates should reflect the actual costs of an efficient operator, and equivalent wholesale broadband access products should be offered for each of the incumbent's retail services.

In short, traditional interconnection regulation has attempted to identify the benefits created by incumbency or dominance and then to replicate those benefits in the interconnection offers made to other operators. In particular, it has sought to even up the contest by offering new entrants access to the economies of scale available to the established operator. This is an almost impossible task, especially as the complexity of the market has increased as the Second Law said it must. One such example is described in Box 2.

Box 2: How the "club effect" has foiled cost-based regulation of mobile termination rates

Cost-based mobile termination rates have generally been set by regulators with reference to the costs of an efficient mobile network operator. The European Commission recommends that the rates are the same for all operators (symmetrical), based on a bottom-up long run incremental cost (LRIC) model, and assuming an operator has a market share of 1/n, where n is the number of mobile network operators in the market. These rates were intended to ensure that all operators had access to the same economies of scale, achievable by all of them if they were equally efficient. In many countries a short period of asymmetry, in which the higher costs of new entrants were reflected in higher termination rates, preceded the imposition of symmetrical rates, so as to counter the legacy benefits of incumbency.

In practice, asymmetrical cost-based rates have failed to counteract the scale advantages of the incumbent operators. For example, in South Africa asymmetry was introduced on a temporary basis in 2011 with the plan for to be phased out in 2013, after it had helped the new entrants and smaller operators to become more established. The level of asymmetry was set initially for one year at 20%, then dropped to 15% and finally, in the last year, to 10%. However, the asymmetry has made virtually no difference to the operator market shares, with the result that the smaller operators are now lobbying the regulator ICASA to maintain the asymmetry a bit longer. Why did asymmetry not work? And why, in other markets, has efficient-operator cost-based symmetry failed to work much better?

The problem has come to be known as the "club effect". All else being equal, subscribers favour the largest operator because they have more on-net call options with this network. On-net prices of a larger network will tend to be lower because of economies of scale, but also because the rational pricing strategy for this operator is to accept lower margins for on-net calls while increasing the price of off-net calls. The smaller operator cannot compete with this strategy. It can lower

[21] In many countries the obligation to interconnect is applied to all operators, not just those with SMP. This provides an additional safety net to ensure that the principle of any-to-any connectivity is achieved, however it is not strictly necessary so long as all operators can interconnect via the incumbent or SMP operator.

its on-net rates but this would not be as powerful a move because there are fewer call options on the smaller network. Equally it may lower its off-net rates to match the on-net rate of the larger operator, which will give its subscribers the same range of discounted calling opportunities as subscribers on the larger network. However, to do this the operator will likely incur a cash-loss because the mobile termination rate is generally higher than the on-net retail charge.

The "club effect" exacerbates any differential in market share between operators. The propensity to make on-net calls is often 2-3 times higher than off-net calls, so there is a major advantage in having more subscribers on the network. Cost-based termination rates on their own cannot resolve this problem: either the asymmetry has to be greater than that justified by cost differentials, or the symmetrical rates need to be so low (ideally zero) so that smaller network operators are able to respond to the on-net price discounts made by the larger players.

The Kenyan regulator CCK's approach has been to regulate the differential between the on-net and off-net retail prices charged by the dominant operator as well as termination rates.[22] A similar approach has recently been adopted in Nigeria where the regulator has outlawed any variation between on-net and off-net retail prices[23]. The European Commission's latest approach of very low "pure LRIC" pricing is an alternative and pragmatic attempt to resolve this problem within the constraints of a cost-based interconnection system. Another possible solution would be to abandon cost-based interconnection altogether in favour of bill-and-keep. Either way mobile termination rates should fall close to zero as fast as possible.

Source: Author

1.4.2 Market power in the Internet

Market power within the Internet domain is different from that within circuit-switched telecommunications. It has also changed significantly through the short life of the Internet. Initially the major backbone "Tier 1" network providers seemed the most powerful players, as they controlled the essential infrastructure and were able to establish favourable interconnection terms through peering among themselves and selling transit to Tier 2 ISPs. An over-supply of capacity and the trends towards regional caching and peering has reduced the influence of these players on a global scale, but they retain significance within specific Internet regions. For example, in many countries there remain only one or two Tier 1 companies[24], with the result that they continue to enjoy market power: e.g. BT in the UK, Deutsche Telekom in Germany, and China Telecom and China Unicom in China.

More recently market power appears to be vested increasingly in the Content Distribution Networks. CDNs are not networks in the traditional sense as they purchase connectivity between their servers, but equally they use this infrastructure to offer transmission services to their customers (the CAPs) in addition to the core business of storing content on their servers. CDNs therefore straddle the divide between content and network, which has provided a conundrum for the authorities as to whether or how they can be regulated. Scale is key to the operations of CDNs as they have to access large quantities of content and locate their servers sufficiently close to users as to enable regional peering relationships to form. So CDNs have increasing market power but many of the companies are hidden from public view (e.g. Limelight, Akamai, Panther, BitGravity, Highwinds) because they provide relatively few retail services.

[22] Interconnection Determination no.2 of 2010 at www.cck.go.ke/regulations/downloads/interconnection_determination_no2_2010.pdf, and the addendum thereto at www.cck.go.ke/regulations/downloads/determination_sms_2012.pdf

[23] Determination of Dominance in Selected Markets in Nigeria, Nigeria Communications Commission, 25 April 2013 at http://ncc.gov.ng/index.php?option=com_content&view=article&id=1029:determinations&catid=66:cat-web-legal&Itemid=210

[24] Defined as an ISP which has access to an entire Internet Region solely via free and reciprocal peering agreements.

There is much dynamism in the CDN sector of the market however one trend in particular may give regulators cause for concern: some of the largest CAPs (e.g. Google, Netflix, Amazon) are increasingly establishing themselves as CDNs, creating the possibility that they will leverage their retail market power within the infrastructure domain. Google in particular has attracted the attention of competition authorities in Europe and the US, both in relation to its dominant position as an Internet search-engine and with regard to net-neutrality.

1.4.3 Can dominance be fully overcome?

Within the shifting-sands of the Internet no single dominant player has emerged, and market power has ebbed and flowed between various players. Even now as concerns mount about the dominance of the large CAP/CDN companies, the market is likely to move on and a new generation of competitors will emerge to contest the market. If the history of the Internet has taught us anything it should be that dynamic markets find their own equilibrium[25] and there is seldom a need for regulatory intervention. The Third Law prevails.

But what of the impact on the wider telecommunications domain? Can dominance be creatively destroyed in this sphere also? The answer is likely to be different for the core network (the legacy circuit-switched network and replacement IP networks) and the access network (the “last-mile” connection to customers that is required regardless of the core-network functionality).

The core network

In the core (backbone) network convergence on IP technologies is already assured. While legacy circuit-switched networks will continue to exist for some time, no new investment in these technologies will be made and assets will be retired progressively as they reach the end of their economic lives. The pace of change will vary by country, depending both on the reach of the legacy network and the degree of economic development. Developed countries generally have a much greater installed base of TDM equipment that will take many years to migrate to IP unless operators are willing to write-off significant past investments. In contrast, some developing countries already operate fully IP core networks[26] because they had so little legacy infrastructure to replace.

The likelihood is that managed IP networks in the future will be increasingly deployed as an overlay on the public Internet for specific applications, and not seek to compete with the public Internet head-on. Today, even as applications become more and more sensitive to latency effects, and are thus natural candidates for a managed IP network, they are almost entirely carried on the public Internet. The rate of improvement in quality of service on the Internet is sufficient for all but the most mission-critical corporate applications with the result that managed IP networks are destined to remain a minor part of the market. In these circumstances, it is probable that stand-alone NGNs will gradually decline as reliance on the public Internet, with a quality of service overlay, becomes the standard networking approach.

[25] Or, more accurately, a sequence of equilibria: the dynamism of the market ensures that it never reaches a final resting position as would be the case in those oligopolistic markets in which the participants cooperate rather than compete.

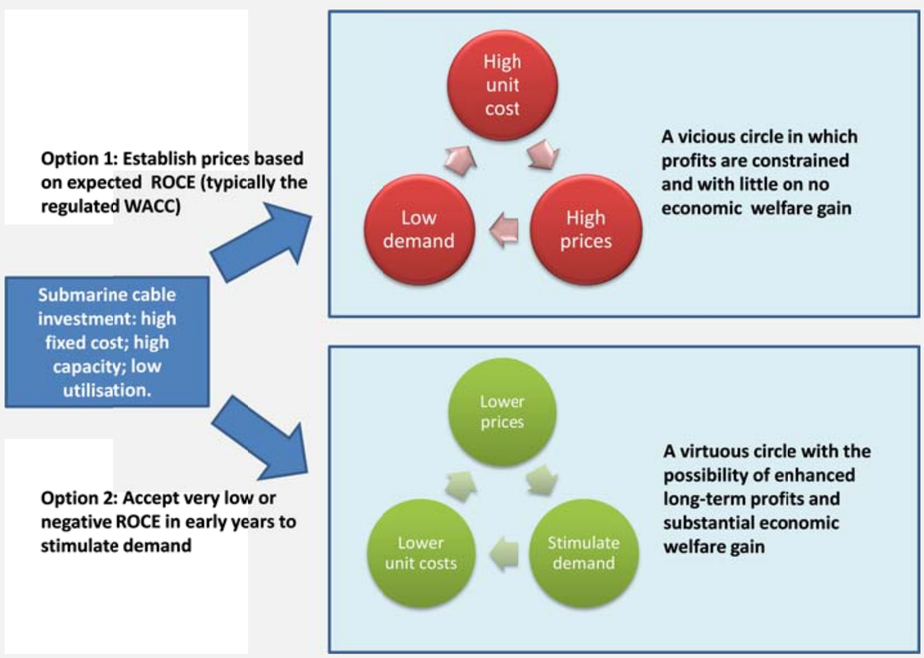
[26] St Vincent and the Grenadines claimed to be the first such country in 2010: http://www.time4lime.com/vc/news/press_release.jsp?view=RG.89

As core networks converge on the public Internet, so the extent of competition will increase. Although there may remain relatively few Tier 1 operators in some countries, they will face strong competition from Tier 2 providers and CDNs will have considerable countervailing buyer power when it comes to negotiating interconnection agreements. Already in many countries there is little or no ex-ante regulation of the core network at either wholesale or retail level. However, in developing countries this situation may still be many years away.

The access network

All services and applications rely ultimately on customer access. Being able to access customers with sufficient bandwidth, with sufficient quality and reliability and at an affordable price is essential to achieve market penetration. Provide these conditions and the applications will follow as broadband history has shown. However, this is a never-ending tension at work and there is just as much possibility of a vicious circle developing as a virtuous one. Figure 7 describes this conundrum in relation to international submarine cable access, one of the main challenges currently facing developing countries as they seek to expand broadband and Internet connectivity.

Figure 7: The relationship between demand, cost and price



Source: Author

The on-going challenge is for the access network to provide sufficient bandwidth to support all the applications that users demand. Price can be used as a tool to curtail demand so as to keep it within the evolving capacity limits of the access technologies. But price is a crude tool, and demand is influenced by many other factors as well so that peak-load is never fully predictable, with the result that some throttling of demand always seems to be required. This is where net neutrality comes in.

Net neutrality is the principle that all electronic communication passing through a network is treated equally – independent of content, application, service, device, sender or recipient.[27] It is interesting to note that this principle is very close to the best-efforts paradigm that drives the Internet: it is not to say that all demand will be met, but that any degradation in service levels will be randomly distributed. What is considered pernicious is when one application or service provider is given priority over another one, as this creates the ability to distort the market. As the access provider is in most cases an incumbent telecoms operator with a position of SMP in the national market for the provision of fixed broadband access, this is a real possibility. However, while BEREC reports some blocking of over-the-top VoIP in Europe, it also finds that incidents have been few and they have been resolved relatively quickly without the intervention of regulators.

It increasingly appears that the concerns about net neutrality are being resolved by the market. There is a growing realisation that access providers and content providers are in a symbiotic relationship – each feeds off the other, requiring the other to help it generate revenues and profits. In economic terms, the presence of increasingly powerful CDNs is providing the necessary countervailing buyer power to curb the dominance of incumbent access providers in national markets. The case of Orange and Google described in Box 3 is a good example.

Box 3: The case of Google and France Telecom (Orange)

In January 2103, Stephane Richard the CEO of France Telecom (Orange) caused a stir by announcing that the company had reached an agreement with Google, in which the latter paid Orange for the delivery of Internet traffic. He claimed that a "balance of forces" was at work: Google provides access to a rich array of content for Orange's customers while Orange provides Google with access to a wide range of customer in Europe and Africa. He claimed that 50% of all traffic on the France Telecom network originated with Google, primarily video applications such as YouTube, itself a Google-owned company. The fees involved in the deal were not disclosed.

The agreement between Orange and Google came after a ruling by the French Competition Authority[28] that said that Orange "may ask to be remunerated for opening additional capacity but it must clarify the commercial and billing relationship between its Internet access and Internet transit businesses" so as to reduce the possibility of margin squeeze affecting smaller ISPs that rely on Orange for transit services. The smaller ISPs, most prominently Free, the fastest growing ISP and hence the one with the greatest capacity problem, restricted access to YouTube and other video sites, and blocked advertisements so as to preserve capacity for other applications. With these smaller ISPs there is clearly no "balance of forces" but the hope is that the peering agreement between Orange and Google, coupled with the regulations on Orange's wholesale business, will ensure that at least some of the benefits are passed onto smaller ISPs.

Source: Author

In a sense these developments are a fulfilment of the Third Law: the power of the former-monopoly fixed-line telecom operator is waning. But it has not gone away completely, and quite possibly never will. The cost of building fibre access networks is sufficiently large that competitive supply is almost always uneconomic. Some countries (e.g. Australia, Singapore) are investing huge sums in developing national broadband networks with ubiquitous fibre access, but typically these networks will themselves become a monopoly incorporating the legacy assets of the incumbent. Market power in the access network will, if anything, become more pronounced as the scale of the required broadband investment increases. It may be held in check by the equally rapid growth of some CDNs, so that they can negotiate interconnection as a balance of equals; but the vast majority of players in

[27] This definition comes from BEREC, but the concept is widely adopted.

[28] The ruling can be found at: http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=417&id_article=1970

the IP world will not enjoy the scale of the CDNs, so there will be an on-going need for regulatory intervention to ensure that they can achieve fair and reasonable access and interconnection terms[29].

1.5 The impact of the Laws on national interconnection regimes

The laws of thermodynamics are not rules that must be followed, but descriptions of underlying principles that collectively define a complex system. Similarly, the laws of IP interconnection do not determine how regulators *must* behave, but collectively they define how regulators *should* act in the interests of efficiency and effectiveness in the complex world of IP networks. Like the laws of thermodynamics they are based on descriptions, only in this case of economic relationships that will be triggered under certain conditions. As a general rule they point to a future in which regulators play much more of a watchful game, observing the market and readying themselves to intervene only when and where market failure arises. This effectively means moving into the space currently occupied by generic competition authorities in many countries. With regards to interconnection as to other potential areas of regulation they should remain largely passive, with interventions being carefully targeted.

In the context of the Internet this approach may seem obvious and non-controversial. But for many regulators it will or could be revolutionary; for all it will be difficult. Interconnection has consumed a disproportionate amount of regulatory effort on circuit-switched networks over the past 20 years, and within that broad envelope, the key focus has been on the establishment of cost-based prices. Such work is valuable as part of a regulatory strategy to prevent anti-competitive practices and develop effective competition; but not always has the regulatory investment yielded an adequate return on capital. Is it time to begin to learn the lessons of the largely unregulated Internet?

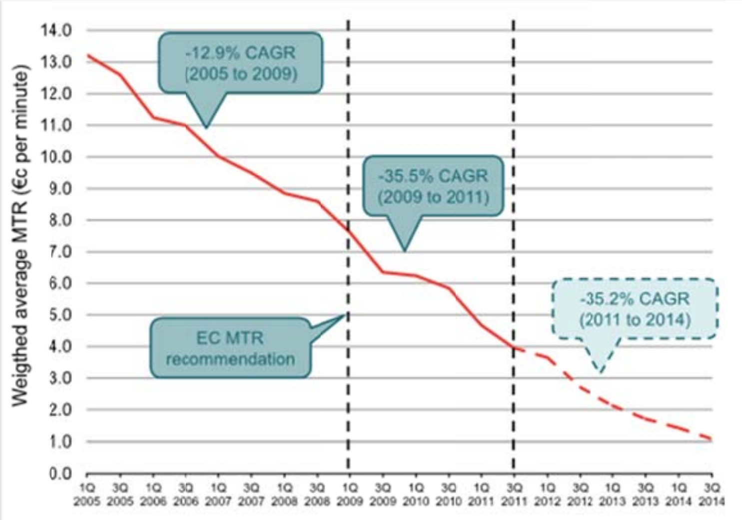
1.5.1 Time to leave cost-based interconnect behind?

In truth those lessons have already begun to be learned and implemented. There has been an almost relentless trend towards lower termination rates, especially for mobile networks. Ten years ago interconnection payments amounted to around 30% of revenue for a mobile network operator, but that figure has dropped substantially as termination rates have fallen. A recent report from the European Union[30] gives a representative picture which is summarised See Figure 8. The decline in fixed termination rates has been considerably less marked, and in some countries rates have actually gone up, but in many cases this has meant that fixed and mobile termination rates are now set at or close to the same level.

[29] See Rogerson, D, "Open Access Regulation in the Digital Economy", ITU, GSR2011.

[30] The impact of recent cuts in mobile termination rates across Europe, A report by Frontier Economics prepared for Vodafone Group: http://www.vodafone.com/content/dam/vodafone/about/public_policy/articles/mtr_impact_of_ec_recommendation.pdf

Figure 8: Mobile termination rates have fallen rapidly over the last 10 years



Source: Frontier Economics

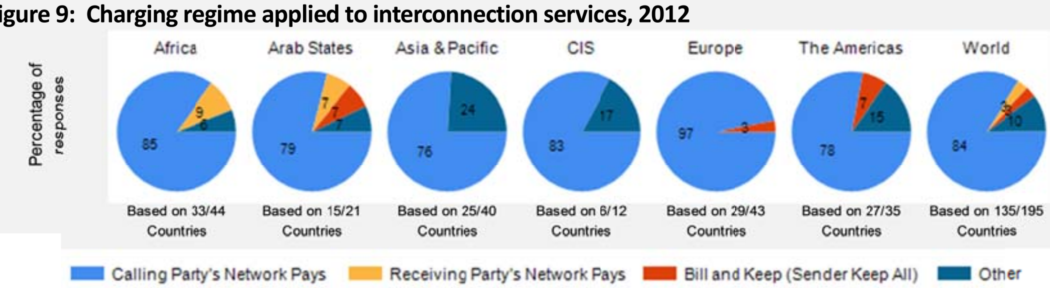
This is not just the result of lower unit costs. It is also the result of a determined effort by many regulators, led by the European Commission (EC), to reduce the effective floor that termination rates seem to establish for retail pricing. Actually the floor relates to off-net retail prices. It is arguable that operators tend to price off-net calls as A + B + C, where A = marked-up network cost of calls, B = termination charge payable to other operator, and C = retail costs and mark-ups. What they should have been doing is reducing B to reflect the net payment after accounting for termination payments received from other operators. The trends towards flat-rate “big bucket” retail tariff plans that give no incremental revenues based on call volumes is now helping operators to reach this outcome, but regulators can assist them by forcing termination rates down towards zero.

Regulators have, however, felt the need to justify the reduction in termination rates in terms of costs. The EC, for example, is forcing mobile termination rates down to under 1 Euro cent per minute by establishing the concept of “pure long run incremental cost” in which the only allowable costs are those which would hypothetically be avoided if the termination service were withdrawn. This excludes all fixed costs and common cost mark-ups, and ensures that economy of scale benefits are fully incorporated in termination rates.

The efforts of the EC in this respect are also proving valuable in overcoming the “club effect” (described earlier in Box 2). However, they could equally have been achieved by withdrawing the requirements in legislation for interconnection rates to be cost-based. This was the approach followed in Hong Kong, China where the regulator OFTA (now OFCA) withdrew its cost-based guidance for fixed-to-mobile and mobile-to-fixed calls, with the result that the operators introduced a bill-and-keep arrangement in 2010. This worked because a clear any-to-any connectivity requirement was in place for voice communications which meant that, absent specific regulatory guidance on pricing, bill-and-keep was the rational default charging policy for the operators.

Although still very much a minority approach (see Figure 9), there are sound economic reasons for favouring bill-and-keep. Not only does it minimise the transaction costs associated with interconnection, but it also ensures that the retail service provider can account for price elasticity of demand without the constraints imposed by cost-based termination. However in some situations paid peering may be more appropriate, especially where there is a marked asymmetry of traffic. For example, downloading a 3-minute video clip uses 35 times as much downstream as upstream capacity, so a network with more content providers will originate far more traffic than a network with more subscribers (eyeballs). The commercial arrangement for peering needs to reflect these different value propositions, which means that it would be inappropriate for regulators to enforce bill-and-keep in all situations.

Figure 9: Charging regime applied to interconnection services, 2012



Source: ITU Tariff Policies Database (http://itu.int/icteye)

But can the price outcome be left to market forces alone? The answer, in most situations, is probably "yes", even though this goes against the grain of regulatory history and practice. The range of factors that establish the value on either side of an interconnection agreement are too numerous and often too nebulous to be calculated with any confidence by the regulator: reach, quality, content, traffic, technology, cost-base all play a part in determining the value balance. Regulators therefore need to take a step back from ex-ante rules and work more from a set of principles that can guide commercial negotiations and, if necessary, be used to establish ex-post resolution to disputes. These principles may include:

- Acceptance that there might be multiple interconnect models in operation at once – depending on the deal between each interconnecting pair of operators.
- For any interconnection agreement that does not deploy bill and keep (settlement free peering) the basis of the charges must be specified and the level of the charges justified in a transparent manner.
- All such interconnection agreements must be equally available to all operators through a transparent Reference Interconnection Offer process. There would be a menu of offers if multiple agreements are in place, so as to ensure non-discrimination.
- Where disputes arise and the regulatory authority has to mediate, interconnection rates should be based on the principle of maximising economic welfare, which in most circumstances will mean setting cost-based rates, and using bottom-up LRIC modelling would be preferable.

1.5.2 Practice forbearance (with one big exception)

The future role of the regulator is going to be much more one of monitoring interconnection agreements rather than intervening to establish interconnection charges. This role is well established with regard to the Internet and, perhaps more by default than design, regulators have adopted a similar approach to managed IP networks (NGNs). However, it is behaviour that will have to be learned in relation to circuit-switched networks, with current regulations gradually being withdrawn as they become redundant. As BEREC warns at the end of its review of IP interconnection, "any measure could potentially be harmful so it should be carefully considered".

In countries where liberalisation is well advanced, significant regulatory "infrastructure" has usually been established to support the goal of effective competition, with particular emphasis on interconnection regulations and cost-modelling. In such situations, it is the danger of stranded regulatory investment that now threatens to stall progress: regulators may be unwilling to unwind procedures that have taken considerable effort to develop, with the result that they rely on regulated, cost-based interconnection rates for too long. In reality, investment in cost models will never be wasted – see Box 4 for an example of good practice – but the way that they should be used is increasingly as regulatory tools to apply ex-post in order to resolve disputes and restrain anti-competitive behaviour. Having such tools available is critical to obtaining a rapid resolution of interconnection problems; and the knowledge that a rapid resolution is possible is critical to ensuring that commercial negotiations take place in good-faith.

Box 4: The future role of cost models in ex-post regulation

The Mozambican regulator, INCM, has recently invested in a range of telecommunications cost models all based on bottom-up LRIC costing principles. Two of them, concerning fixed and mobile termination rates, were constructed despite the fact that the network operators had reached commercial agreement on symmetrical rates. The other three models, concerning rural infrastructure sharing and access to capacity on national backbone and international submarine cables, were constructed despite the fact that there had been no regulatory finding of dominance in these markets to justify the imposition of cost-based rates nor were any regulatory disputes pending. Nevertheless, cost models are seen as a powerful regulatory tool to keep INCM well-informed about market developments. They are being used in scenario planning to anticipate potential regulatory pressure points and to enable the regulator to prepare its response, should competition on its own not prove able to deliver national economic policy goals.

Source: INCM

Ex-post regulation of interconnection will increasingly become the norm, but broadband access is one area that will continue to need cost-based ex-ante regulation. In every economy, and especially in developing countries with relatively poor fixed network access infrastructure, the major bottleneck for the digital economy is in the availability, capacity and cost of broadband access. Duplication of access infrastructure is frequently uneconomic (as is sometime the case for backhaul facilities) and especially so when major new investment is required to achieve broadband access speeds. Regulators therefore need to adopt measures that will encourage and reward investment while retaining open access. Prices should ensure that fixed network operators receive a fair return on their investment, i.e. a weighted average cost of capital that is sufficient, and sufficiently secure, as to encourage them to make the necessary broadband infrastructure investments. Equally service providers need to know that they are able to access this new infrastructure on terms that are fair and reasonable, i.e. equivalent to those of the network operator should it be competing in the relevant services market.

Detailed proposals to achieve this balance have recently been published for consultation by the European Commission[31] and they may serve as guidance for other jurisdictions as well. The key principles are that:

- A national broadband policy[32] should establish the target capacity and reach of broadband next generation access (NGA) networks
- Prices should be established on the basis of a bottom-up LRIC methodology, establishing the current cost of a hypothetical efficient operator using an NGA capable of meeting the broadband policy objectives
- The regulatory asset base for the model should be valued at full replacement cost except for existing civil engineering assets that are capable of hosting the NGA which should be valued on the basis of inflation-adjusted net book value
- The price of access products provided over copper networks should be reduced from NGA levels to reflect lower performance and cost-savings relative to the NGA
- Effective non-discrimination is best achieved through an Equivalence of Input (EOI)[33]
- requirement, so that alternative service providers can replicate the service offerings of the broadband access network operator
- To be effective EOI needs to incorporate service level agreements and key performance indicators, covering matters such as:
    - Ordering process
    - Provision of service
    - Quality of service
    - Fault repair times
    - Migration.

These regulatory principles have widespread application. In Europe the focus of regulatory efforts is to ensure fair and effective competition among service providers which rely on access to an incumbent operator's infrastructure. In developing economies regulatory focus may instead be on ensuring that the incumbent network operator receives a fair payment from much larger and financially stronger service providers seeking to utilise its network to access customers. In relatively few cases will equilibrium, such as that described between Orange and Google in Box 3, be achievable absent regulatory intervention. The EC proposals provide an approach that should be applicable in all other cases.

Conclusions

The emergence of the Internet has radically changed the technology approaches, market philosophies and regulatory paradigm of the telecommunications industry. It has initiated a series of challenges with which telecoms regulation is still coming to terms. This paper has argued that, in

[31] See: http://ec.europa.eu/digital-agenda/en/news/draft-commission-recommendation-consistent-non-discrimination-obligations-and-costing

[32] In the EU this is the Digital Agenda for Europe, which has set the objective of basic broadband for all by 2013 and next generation networks (30Mbps) for all by 2020.

[33] Equivalence of Input means that in all respects (e.g. price, functionality and delivery time) the service received by an independent service provider is the equivalent to that received by the retail division of the network operator.

relation to interconnection, the fusion of Internet and telecoms regulation for the advancement of the digital economy can best be achieved through the following regulatory principles:

1. Set an overall principle of any-to-any connectivity but allow the market to establish how interconnection is performed in order to achieve it.
2. Do not extend regulation from circuit-switched networks to IP networks unless it is proved to be justified and proportionate.
3. Keep interconnection regulation as simple as possible to avoid unintended consequences.
    a. Establish "bill and keep" or "free peering" wherever possible
    b. If termination charges continue be regulated, bring them down towards zero as fast as possible
    c. Do not mandate minimum quality of service standards, other than those which apply to circuit-switched voice telephony
    d. Increasingly focus on principles of transparency and non-discrimination.
4. Regulate primarily on an ex-post basis. Use existing regulatory tools such as cost models to inform decisions that may have to be taken ex-post (e.g. on equivalence of access and margin squeezes) and to ensure that such decisions can be taken rapidly and effectively.
5. Retain ex-ante cost-based regulation for wholesale broadband access (and backhaul in remote and rural areas) to ensure both that there are sufficient investment incentives for next generation access technologies and open access for competing service providers.