wik-Consult • Report

Study for the Federal Network Agency (BNetzA)

# Framework for Interconnection of IP-Based Networks – Accounting Systems and Interconnection Regimes in the USA and the UK

Author:

J. Scott Marcus

Bad Honnef, March 27, 2006

wik CONSULT

# Contents

## List of Tables



## List of Figures

# 1 Introduction

In the United States and the United Kingdom, networks are rapidly evolving to reflect new design principles based on the Internet Protocol (IP) and Voice over IP (VoIP), an evolution that can be viewed as a manifestation of *convergence*. This technological transformation has profound implications in terms of market structure and also in terms of regulation.

Both countries are technologically advanced, and both countries have tended to break new ground in their approaches to regulation (and in some cases to lack of regulation). Both countries have tended to advocate a market-based approach, with as little regulation as possible.

Despite these obvious similarities, there are very substantial differences in approach between the US and the UK. The UK operates within the European Regulatory Framework for electronic communications, which places a substantial emphasis on addressing such Significant Market Power as may exist. The US by contrast often seems bent on deregulation even where competition is ineffective. Also, the US emphasis on Bill and Keep (the absence of regulatorily mandated call termination charges) has worked well both for mobile telephony and for the Internet.

What trends can we identify in the approach that the US and the UK are taking to interconnection in a new, converged world? How are their approaches similar, and how different? What lessons can the BNetzA learn from studying US and UK approaches to interconnection in an IP-based world?

Section 2 describes UK developments, including BT's planned implementation of a Twenty-First Century Network (21CN), Ofcom's new arrangements with BT, and the various consultations Ofcom has conducted and is in the process of conducting. Section 3 describes developments in the very different environment of the United States. Section 4 provides brief conclusions and recommendations.

# 2 Developments in the UK

Section 2.1 deals with BT's recently announced plans for a Twenty-First Century Network (21CN), while Section 2.2 describes the recent arrangements agreed between Ofcom and BT to assure Equality of Input to alternative network operators. These sections are provided for completeness, but are intentionally kept at a cursory level since most readers will already be familiar with these developments.

Section 2.3 covers Ofcom's pertinent regulatory proceedings in some depth. Section 2.4 notes additional developments in the UK, particularly the emergence of settlement-free IP interconnect among VoIP providers.

## 2.1 BT's 21CN

British Telecom (BT) has announced an ambitious migration to a 21st Century Network (21CN). The 21CN is a single IP and DWDM-based network that will carry both voice and data.[1]

BT is hoping that this evolution will enable them to (1) transform the customer experience, (2) accelerate time-to-market for new services, and (3) eliminate about a billion pounds per year in operating expense.

It is the simplification of BT's network, and the corresponding elimination of redundant function, that would drive the cost savings. BT's plethora of legacy networks are supported today by some 3,000 Operational Support Systems (OSS). Consolidation into a single integrated network promises a huge reduction in the IT costs associated with these systems, and also the elimination of as many as 100,000 network devices.[2]

At a technology level, 21CN does not appear to be particularly novel or radical. Integrated voice and data networks have been commonplace for many years. 21CN is to based on technologies that have long been available and stable – DWDM, DiffServ, MPLS traffic engineering, and VoIP.

Nonetheless, 21CN is potentially an important and radical departure. What is distinctive is BT's apparent commitment to migrate their entire network from classic PSTN telephony to exclusive use of VoIP; their willingness to commit to specific migration objectives in a comparatively short period of time; and particularly their willingness to commit their regulatory future to this course of action.

---

1 See http://www.btglobalservices.com/business/global/en/business/business_innovations/issue_02/century_network.html.

2 Remarks of BT CTO Matt Bross at the "NGN and Emerging Markets" workshop, December 5, 2005.

The migration is supposed to be 50% complete in 2009. Given the size and complexity of BT's network, this is an ambitious migration target. At the same time, the risks to the schedule are likely to rest, not in the challenges of deploying the technology, but rather in (1) the difficulty of migrating customers from legacy offerings, and (2) the complexities of migrating Operational Support Systems from the old world to the new.

The 21CN initiative relates to the core of the network. BT is also committed to substantial upgrades at the access level. In a regulatory sense, these are addressed through the recent agreement between Ofcom and BT, covered in the next section.

## 2.2 The Ofcom / BT agreement

In June 2005, Ofcom announced an agreement with BT that represents a significant departure from previous regulatory practice.[3] BT made legally enforceable commitments[4] to provide a range of access services to competitors on a nondiscriminatory *equivalence of input* basis. Ofcom defines *equivalence of input (EoI)* as "…a requirement for BT to make available the same SMP products and services to others as it makes available to itself, at the same price, and using the same systems and processes." EoI obligations would be applicable "… when the cost is proportionate, and in particular [to] all new wholesale SMP products, processes and systems, and therefore to all new SMP products delivered over 21CN."[5]

BT has agreed to make key wholesale offerings where it has previously been found to possess SMP available to competitors *on an equivalence of input basis*. Most if not all of these are, to be sure, available today in connection with *ex ante* remedies imposed in response to SMP. What the commitment implies is that these wholesale services must as of these dates be delivered by BT's Access Services Division (Openreach) using new order processes and Operational Support Systems. The target dates are:

- LLU – ready for service June 2006
- WLR on the PSTN – ready for service mid 2007, migration complete June 2010
- WLR on ISDN2 - ready for service September 2007, migration complete end-March 2009
- WLR on ISDN30 – ready for service December 2007, migration complete December 2009

---

**3** See http://www.ofcom.org.uk/media/news/2005/06/nr_20050623 and http://www.ofcom.org.uk/consult/condocs/telecoms_p2/statement/main.pdf. See also Ofcom's *Final statements on the Strategic Review of Telecommunications, and undertakings in lieu of a reference under the Enterprise Act 2002* (Strategic Review), 22 September 2005.

**4** BT offered undertakings in lieu of a reference by Ofcom under the Enterprise Act. The undertakings are thus pursuant to competition law, and operate in a parallel and complementary fashion to Ofcom's *ex ante* sector-specific regulation. See http://www.ofcom.org.uk/consult/condocs/sec155/sec155.pdf. BT's commitments appear as Annex A to Ofcom's *Strategic Review*.

**5** Further Consultation, op. cit., section 1.21.

- IPStream – ready for service end-December 2005, migration complete end-December 2006
- Wholesale Ethernet Service (WES), and Backhaul Ethernet Service (BES) – ready for service September 2006, migration complete March 2007.

For three additional wholesale products where BT has SMP (Partial Private Circuits, Carrier Preselection, and ATM interconnection), BT has committed to make offerings available that are "… sufficiently comparable to allow competition to take place – so called 'equivalence of outcome'."

BT has not been broken up, but a substantial "Chinese Wall" has been established between BT's new Access Services Division (ASD) (referred to in more recent BT announcements as "Openreach") and the rest of BT. The ASD will have a separate management team with substantial autonomy. It have 30,000 employees, who will over time have their own uniforms and their own branding. Notably, their bonus plans will be based on ASD objectives, and will be decoupled from the price of BT group stock. An Equality of Access Board will monitor ASD's compliance with its commitments to provide equality of access.

## 2.3 Ofcom's consultations

Ofcom has conducted a number of public consultations on the significance of the migration to NGN, and on the impact of that migration on regulation in general and on interconnection in particular. The documents provide a wealth of enlightened and informed analysis; at the same time, relatively little has concretely been implemented to date. To a point, that is as should be: it would have been premature to attempt to design in detail a regulatory regime today for an environment that is still to a significant degree speculative.

Instead, Ofcom has focused on putting in place processes and mechanisms for moving the regulatory environment forward over time, as the migration to 21CN progresses. The focus to date has thus been on *process* rather than on *outcome*.

Among the relevant Ofcom consultations (all available on Ofcom's website) are:

- *Next Generation Networks – Future arrangements for access and interconnection* (First Consultation), October 24, 2004
- *Next Generation Networks: Further consultation* (Further Consultation), June 30, 2005
- *Final statements on the Strategic Review of Telecommunications, and undertakings in lieu of a reference under the Enterprise Act 2002* (Strategic Review), 22 September 2005.

- *Ofcom's approach to risk in the assessment of the cost of capital*, January 26, 2005 (updated February 2)
- *Ofcom's approach to risk in the assessment of the cost of capital: Second consultation in relation to BT's equity beta*, June 23, 2005
- *Ofcom's approach to risk in the assessment of the cost of capital: Final statement* (Final Statement), August 18, 2005
- *Review of BT's network charge controls: Explanatory Statement and Notification of decisions on BT's SMP status and charge controls in narrowband wholesale markets*, August 18, 2005

In the sections that follow, we consider first those that relate to NGN interconnection in general (in Section 2.3.1), and then those that deal more narrowly with cost and cost modeling issues (in Section 2.3.2).

### 2.3.1 Ofcom's consultations on NGN Interconnection

In these proceedings, Ofcom was looking to consider a range of questions. On the one hand, they wanted to understand what they should do about existing SMP obligations during a period of transition; on the other, they wondered what new SMP obligations and offerings might be necessary and appropriate in the new world of the NGN. An overlay on all of these considerations was cost causation and corresponding responsibility: To the extent that BT unilateral decisions stranded the investments of alternative network operators (altnets), how should financial responsibility be apportioned?

Ofcom has tended to view the migration to NGN both as a challenge and as an opportunity. Specifically, they wanted to understand to what degree this migration might present an over-arching opportunity to move decisively but selectively away from *ex ante* sector-specific regulation to primary reliance on *ex post* application of competition law, as has been their goal for some time. This is clearly expressed in the First Consultation[6]:

> " … BT's planned move to 21CN raises many questions and issues for these existing regulated products. All of them will have to evolve if they are to be effective as a means of enabling competition in a world of NGNs. However, this is not an argument for remaining with the status quo. Major technology changes, which occur naturally in competitive as well as regulated markets, are always likely to disrupt existing models of

**6** Section 2.14

> competition. Rather, the move to 21CN should also be viewed as creating the first ever opportunity to ensure that access and interconnection to an incumbent's network supports competition from the outset, thereby creating an environment where regulation can be focused on key bottlenecks and rolled back elsewhere. … "

This can be viewed as being consistent both with their generally deregulatory objectives, and specifically with their goals in their accommodation with BT. They have committed themselves to the following regulatory principles[7]:

1. promote competition at the deepest levels of infrastructure where it will be effective and sustainable;
2. focus regulation to deliver equality of access beyond those levels;
3. as soon as competitive conditions allow, withdraw from regulation at other levels;
4. promote a favourable climate for efficient and timely investment and stimulate innovation, in particular by ensuring a consistent and transparent regulatory approach;
5. accommodate varying regulatory solutions for different products and where appropriate, different geographies;
6. create scope for market entry that could, over time, remove economic bottlenecks; and
7. … unless there are enduring bottlenecks, adopt light-touch economic regulation based on competition law and the promotion of interoperability.

It is worth noting that Ofcom has drawn a distinction between evolutionary developments in the core of the network and those at the edge. "The issues raised by the migration to next generation access networks, ie the migration from copper to fibre based access, are distinct from the issues raised by the migration to NGNs. This consultation does not aim to address these issues. Ofcom is considering them in a separate workstream."[8] For the most part, these NGN access issues have been addressed by the Telecoms Review and by Ofcom's agreement with BT.

---

**7** *First Consultation*, section 1.9.
**8** *Further Consultation*, section 1.4.

#### 2.3.1.1 Old and new SMP offerings

As regards existing SMP obligations, and specifically existing SMP interconnection offerings, they came to the unsurprising conclusion that those offerings would need to be maintained for some period of time. At the same time, they came to the equally unsurprising realization that new SMP interconnection offerings would be appropriate in the future. This necessarily implies some period of overlap:

> To enable business planning for alternative providers there initially needs to be continuity of existing SMP products (those products that BT is obliged to offer in markets where they have Significant Market Power), but we believe that this should only be for an interim period during which both legacy and next generation products are available. To ensure a timely move to next generation interconnect we propose that legacy products should be withdrawn once there is no longer reasonable demand or when next generation products provide an adequate replacement that providers are able to migrate to.[9]

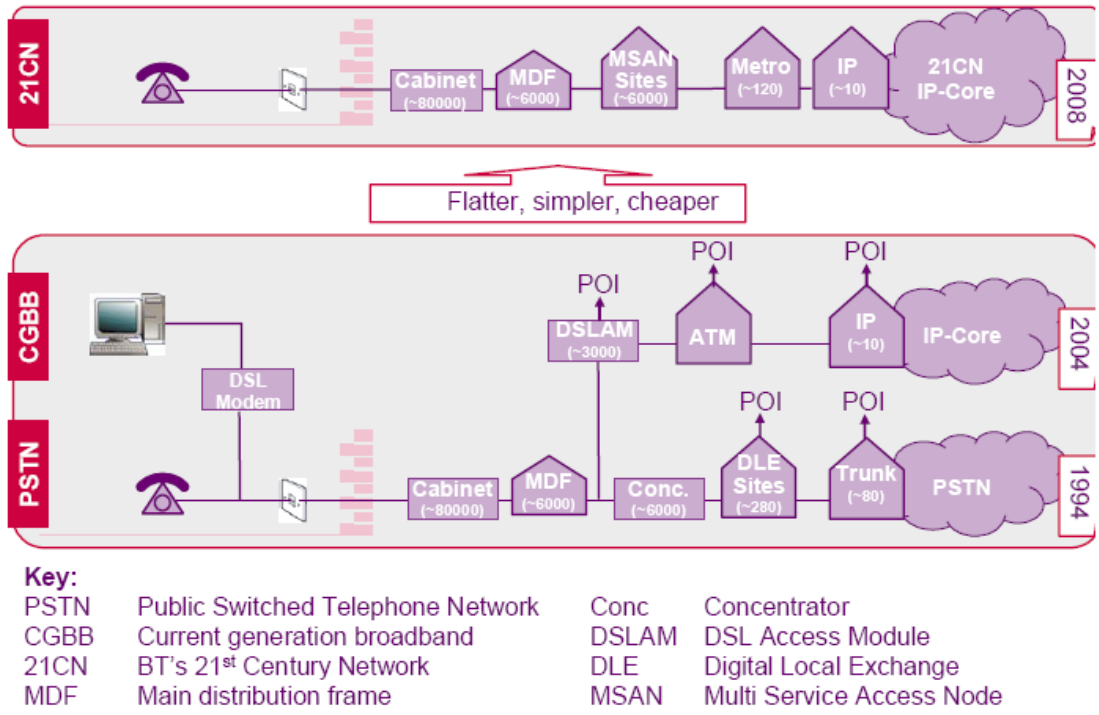
#### 2.3.1.2 Compensation arrangements for SMP product migration

Ofcom recognized[10] that the new structure that BT envisioned for 21CN necessarily implied a flatter network with fewer points of interconnection. Today, BT has some 3,000 locations at which competitors can connect to the DSLAM, and some 280 Digital Loop Exchange (DLE) sites at which competitors can gain access to the voice network. In the 21CN as currently envisioned, interconnection will be possible only at the metro nodes, i.e. only at 100 – 120 sites.[11]

---

**9** Ibid., section 1.11.
**10** *First consultation*, Figure 1.
**11** Ibid., section 2.9. Figure 1 of the *First Consultation* shows 120 sites; however, at the December 5, 2005 "NGN and Emerging Markets" workshop, hosted by WIK on behalf of BNetzA, BT CTO Matt Bross referred to 100 metro sites.

Figure 1: Comparison of existing BT voice and broadband networks with 21CN




Source: Ofcom, *First consultation*, Figure 1, page 11.

This inevitably raised many questions. Alternative operators had invested significant sums to interconnect with BT at existing interconnect locations. Now, as a result of unilateral decisions set in motion by BT, many of those interconnect locations would no longer exist. How should these costs be apportioned?

A particular challenge related to the Multi Service Access Nodes (MSANs), which BT initially envisioned as primitive facilities offering little scope for interconnect. The responses to the First Consultation on this point[12] are illustrative:

> E.20 BT committed to examining the commercial and technical feasibility of MSAN interconnect with the industry, stating that they will ensure that any … decisions will retain the flexibility to offer MSAN interconnection.
>
> E.21 Other respondents stated that it is not clear, at the moment, whether MSAN interconnect is required. Further information is needed including the costs, from

**12** *Further Consultation*, Annex E, pages 12-13.

BT, of the different interconnect options. One alternative provider strongly disagreed with MSAN interconnection as it felt it would undermine the whole LLU business model. Fibrenet stated that "No MSAN interconnect is going to be practical".

E.22 However, there were several supporters of MSAN interconnect, one strongly urging Ofcom to ensure MSAN interconnection, with another stating that providers want to maintain the benefit of built out networks. Vodafone's view was that regulation should be focused on access and interconnect at the MSAN. Energis stated that MSAN interconnection is required, but is unlikely to be viable in areas where LLU is not viable. In general, there were differing views on the technical level of access that should be provided.

E.23 Opposing views were given regarding the control of BT's MSANs for voice access. One alternative provider stated that BT should allow other provider's call servers to control MSANs, whilst another claimed that allowing multiple Altnets to control MSANs was too risky.

Ofcom found[13] "…that the key factors relevant to compensation arrangements for BT's 21CN migration are:

- the extent to which these changes are unilaterally decided by BT without industry agreement;
- the distribution of benefits that accrue from these changes;
- the remaining life of any legacy interconnect equipment employed at the time of the change;
- the extent to which new interconnect investments are made by communication providers after they have been made aware of forthcoming changes that would impact that investment; and
- the additional cost necessarily and directly incurred as a result of having to bring forward investment in new interconnect equipment.

#### 2.3.1.3 The migration process

The Ofcom consultations call for a significant period of overlapping old and new SMP offerings (implying higher costs to BT during a period of parallel operation), and also for

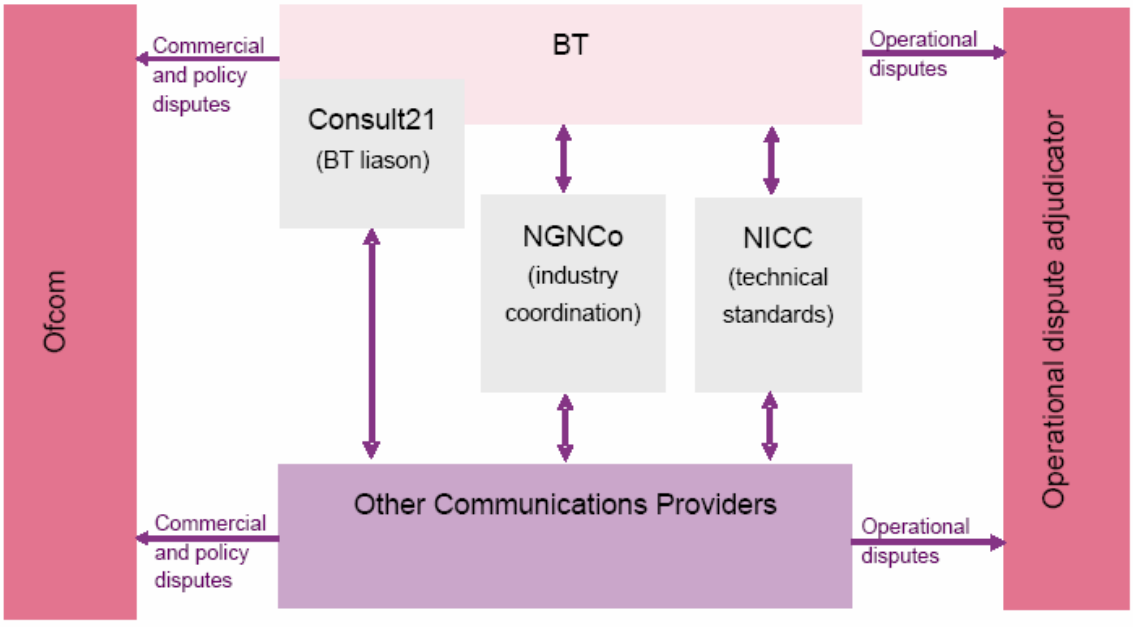
**13** *First Consultation*, section 1.13.

a significant migration planning effort. Planning would be needed at an operational level, and at a technical standards level; moreover, consultation between industry participants would be required, as well as timely information to consumers. To this end, Ofcom has placed considerable emphasis on the mechanisms to enable BT and its wholesale customers/competitors to cooperatively consult and plan for the transition.[14]

These consultation mechanisms can be viewed as reflecting an appropriate emphasis on *process* rather than on *outcome*. Given the degree of uncertainty today as regards the ultimate evolution of 21CN, Ofcom has prudently avoided prematurely locking in specific regulatory policies; instead, it has focused on the establishment of relatively informal institutions that would facilitate industry efforts to arrive at reasonable accommodations as the process moves forward, and has retained the ability to intervene when necessary.

Ofcom envisions the whole process[15] working like this:

Figure 2: Overview of NGN process proposals




Source: Ofcom, *Further Consultation*, Figure 9, page 33.

**14** The discussion in this section is based on the Section 4 of the *Further Consultation*.
**15** Ibid., Figure 9, page 33.

#### 2.3.1.3.1 Consultation bodies

To date, two major fora have been driving the process. The first is *Consult21*, a forum created by BT to facilitate open cooperative discussions with its wholesale customers on the migration of its existing SMP products, and to begin to consider future SMP products as 21CN matures. Consult21 appears to be working reasonably well. As one illustration, Steve Hewson (MCI) recently remarked that BT's openness and transparency in these consultations had been extremely helpful, and that this kind of open dialogue is key to sustained viability and investment.[16]

The second is the *Network Interoperability Consultative Committee (NICC).* The NICC is responsible for technical standardisation of interconnect interfaces within the UK, drawing on the work of other standards bodies (e.g. ETSI, ITU-T, and the IETF). The NICC is currently constituted as an advisor to Ofcom. Ofcom is concerned that the perception of the NICC as an instrument of the regulatory body reduces its effectiveness. Ofcom has announced its intention to transform the NICC into an independent industry-owned body.

Ofcom also proposes to add a third consultation forum, referred to in the *Further Consultation* as *NGNCo*, to the mix. The next section discusses NGNCo in greater detail. Briefly, Ofcom appears to be concerned that Consult21 is too close to BT. NGNCo would be a neutral industry forum, with Ofcom participation, that would be authorized to deal with a number of NGN migration issues, including:

- Producing a reference interconnection architecture.
- Producing a transition plan setting out the detailed process for managing the transition from existing to NGN networks (including BT and other providers' NGNs), including the process for migrating PSTN interconnection to NGN interconnection.
- Producing a communications plan setting out how this transition will be communicated to consumers.
- Overseeing the actual transition, taking any such action as may be necessary in order to ensure that the above plans are achieved; however, the group would not be responsible for managing the deployment by BT or any other communication provider of their NGNs.

Ofcom apparently intends that NGNCo would complement the activities of Consult21 and the NICC; however, there is considerable opportunity for overlap. We return to this point in the next section.

---

**16** Remarks at the "NGN and Emerging Markets" workshop, December 5 2005.

Finally, Ofcom has suggested that some kind of fast track *operational dispute adjudicator* might be able to resolve minor disputes associated with the migration to NGN more quickly and at less expense than can Ofcom itself.[17]

#### 2.3.1.3.2 The NGNCo

Ofcom commissioned a consultant's report on a proposed structured for NGNCo.[18] The resulting document is, in many respects, a well-reasoned and thoughtful document: at the same time, the *Consultants' Report* embodies many mutually contradictory assumptions about how the organization is to function, and how it should achieve its intended purpose.

Key parameters of NGNCo as proposed[19] are:

- The body's purpose should be to develop a joint vision and framework for the transition to NGNs that encompasses commercial, technical and operational issues
- It will issue recommendations to the industry
- In order to function effectively, the body will require a stringent governance and organisational structure:
    - its membership should be inclusive; stakeholders will be able to participate in workgroups on specific issues independent of size and degree of infrastructure ownership
    - in order to function effectively, the NGN body will need a strong figurehead as a chairman supported by a well-resourced management team equipped with strategic, technical and programme management expertise
    - a board consisting of 8-10 industry representatives needs to be appointed to approve the body's recommendations and to drive its agenda forward

---

**17** "Inevitably, there may be circumstances in both the planning and implementation of this change where the parties involved cannot agree. Ofcom's view is that a fast track adjudication scheme for alternative dispute resolution, would be preferable to Ofcom's formal dispute resolution powers for resolving operational disputes. This is because the migration to NGNs is likely to be time critical and because formal dispute resolution under Ofcom's legal powers is likely to be longer and more resource demanding than adjudication." *Further Consultation*, sections 4.39 to 4.43.

**18** Spectrum Strategy Consultants, *Proposal for Discussion: Ofcom: Scoping an NGN industry body (Consultants' Report)*, December 9, 2005.

**19** The following bullets are taken from the Executive Summary of the *Consultants' Report*.

  - the organisation will be independent, accountable only to its members
  - Ofcom should adopt the role of an active observer

- The body should be owned and funded by industry in order to enable its independence. However, Ofcom will need to play a leading role during the set-up phase
- The body is required as soon as possible and should be set up within the next six to seven months; it is expected to have a lifetime of 3-4 years

In this section, we consider in some depth four key aspects of NGNCo as proposed that are likely to influence its effectiveness. These are (1) its mission, (2) its membership, (3) its organizational structure, and (4) its relationship to the other consultative bodies associated with the transition to NGN in the UK.

As regards NGNCo's mission, NGNCo is conceptualized as offering "a neutral ground for negotiation between BT and the rest of the industry", a central point to enable "operators and service providers to exchange views more efficiently", and a means of enahcing transparency as regards the migration of BT and other operators.[20] NGNCo's mission would be primarily of policy (at a level below that of Ofcom), strategic questions, and operational planning (but in this case at a higher level than that of the BT-led *Consult 21* process).

As regards membership, the *Consultants' Report* accepts stakeholder input that the organization needs to be "inclusive but effective", and to this end proposed a tiered membership structure. Membership would be limited to network infrastructure operators, and to organizations that provide end user services over those networks, including ISPs, virtual operators, applications developers, and technology solution providers. Different classes of member should have different obligations and different prerogatives. Board membership, in particular, should be restricted to organizations that are making substantial investments in NGNs.

The *Consultants' Report* envisions an NGNCo that would "… aim to be a decision making body that issues guidelines to the industry".[21] The *Consultants' Report* emphasizes that stakeholders saw no need for an additional debating society; at the same time, the report rejects the notion that NGNCo should make binding commitments on behalf of the industry.

NGNCo is to make policy decisions on a consensus basis. It is to be a private corporation, with no explicit powers other than those implicit in its unique relationship to

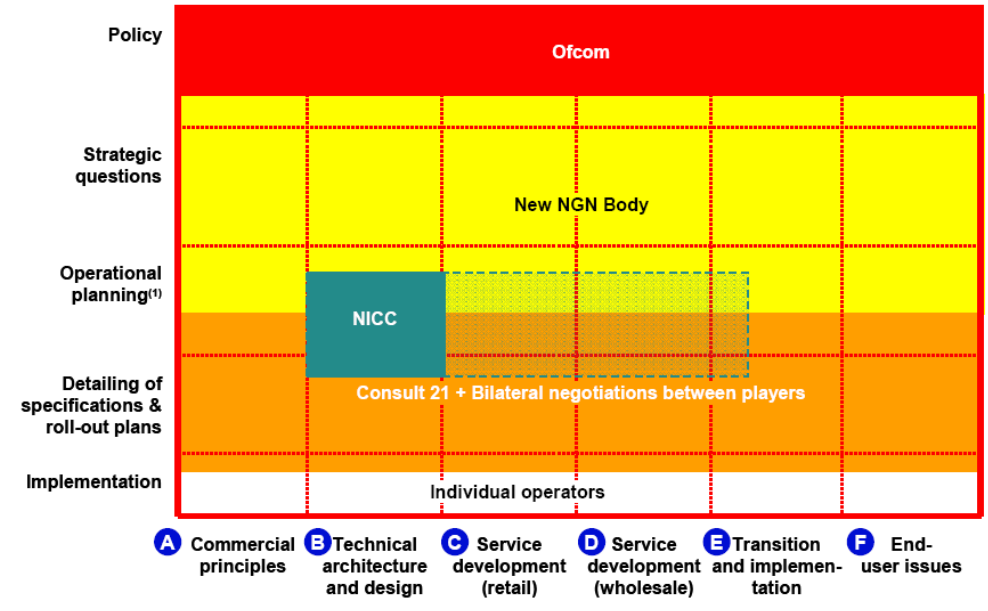
**20** *Consultants' Report*, page 2.
**21** Page 21.

Ofcom. It is to have no enforcement powers. Its decisions should be amenable to some independent appeal process.

It is difficult to see how NGNCo's role as a decision-making body can be reconciled with these assumptions about its powers and structure. It would appear that it will be able to make decisions only in instances where a consensus already existed, or could be forged through negotiation. In the case of interconnection issues – one of its primary explicit areas of activity – these conditions are unlikely to be fulfilled. Inevitably, any change will benefit some parties and harm others. Consensus will prove to be elusive.

Finally, it is worth examining the relationship between the proposed NGNCo and the various other consultation bodies and processes that have a role to play in BT's NGN transition process – Ofcom's own consultation processes, the NICC, and the BT-led Consult 21. As the *Consultants' Report* observes, a number of stakeholders expressed concern about possibly overlapping or conflicting responsibilities. The *Consultants' Report* envisions[22] a matrix of responsibilities:

Figure 3: Position of the proposed NGN Body in the NGN transition process

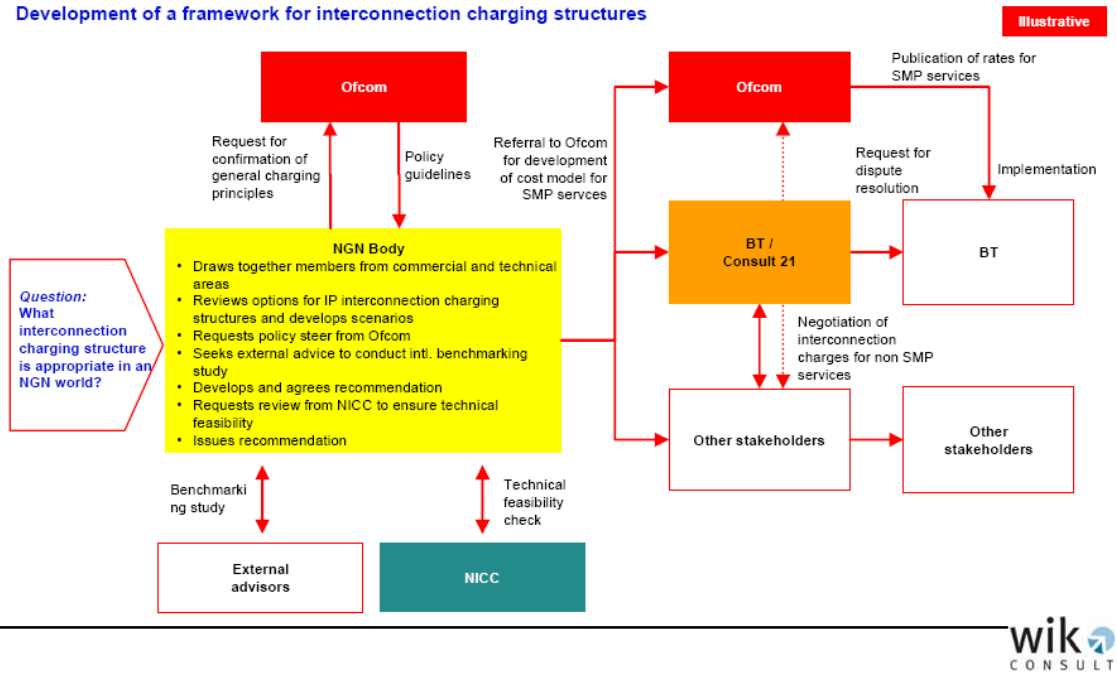



Souce: Spectrum Strategy Consultants, *Proposal for Discussion: Ofcom: Scoping an NGN industry body*, Exhibit 9, page 13.

**22** Page 13.

Clearly, there is a substantial risk of overlap, duplication and friction. In the concrete case of interconnection arrangements, the *Consultants' Report* envisions the following process[23]:

Figure 4: Development of a framework for interconnection charging structures




Source: Spectrum Strategy Consultants, *Proposal for Discussion: Ofcom: Scoping an NGN industry body*, Exhibit 13 (upper portion), page 19.

Ofcom's attempts to establish an open, transparent industry-led process must be applauded; nonetheless, the concern that must be raised as regards the proposed specific arrangements is that they create expectations that NGNCo cannot possibly fulfill. Key concerns include:

- NGNCo cannot simultaneously be expected reach controversial and difficult decisions and to operate by consensus; moreover, its effectiveness in situations where consensus does not exist, and cannot be achieved, will necessarily be limited by its inability to enforce its decisions (unless it can arrange for Ofcom to enforce decisions on its behalf). Ultimately, the hard decisions will have to reached by the organization that has the authority and the processes in place to address them: Ofcom.

**23** Page 19.

- NGNCo cannot simultaneously be inclusive and exclusive; moreover, to the extent exclusive, it would need clear and unambiguous critieria for exclusion. Using the level of contribution as such a criterion is a double-eded sword. Conditioning board memberships on the level of annual contribution[24] is common among trade groups; however, to the extent that NGNCo has quasi-regulatory functions, this practice risks undermining the organization's credibility by creating the impression that influence is up for sale.

- The overlaps in authority and responsibility between NGNCo, Consult 21, the NICC, and Ofcom itself place substantial coordination burdens on all of these organizations and invite turf wars.

- Most generally, NGNCo is being asked to solve a great many distinct shortcomings in the current arrangements in the UK, without being given (under the proposed arrangements) authority commensurate with its proposed responsibilities.

### 2.3.2 Ofcom's consultations on cost modeling

Ofcom recognized several interrelated factors of the current regulatory and market environment that could influence BT's profitability and that of its competitors, including:

- The higher risk to BT and its shareholders in implementing 21CN

- The ultimately lower unit costs of operation for 21CN

- The losses to competitors associated with stranded investments in interconnection facilities

- The cost to BT of simultaneously offering both old and new SMP offerings during the period of transition

There appears to be no explicit discussion in the documents of the losses to BT associated with abandonment of traditional PSTN assets (that were not yet fully depreciated) as a result of the migration to 21CN. Ofcom presumably considers that BT already implicitly factored in these costs into its initial capital expenditures when it voluntarily proposed to migrate to 21CN.

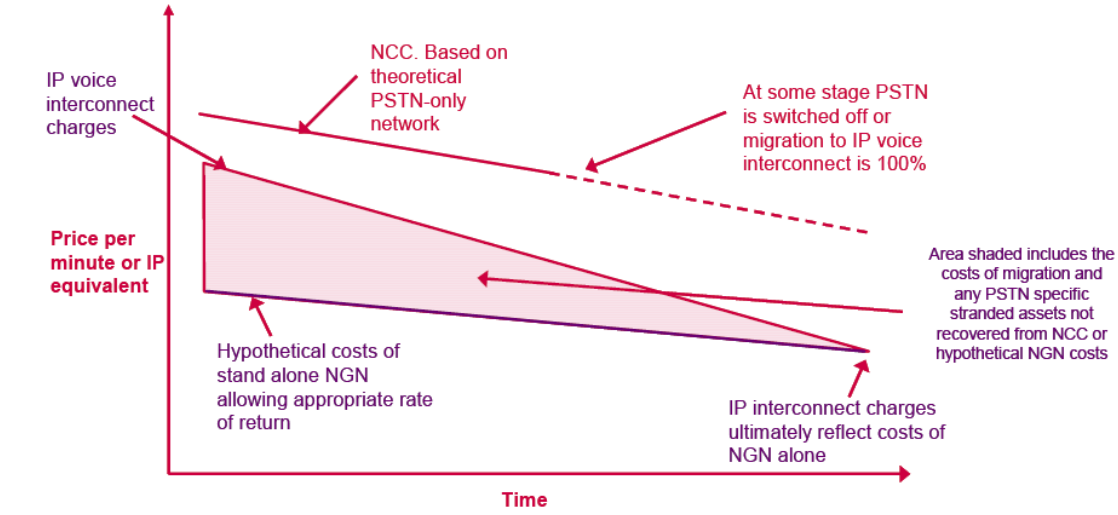
Ofcom has put forward the following over-arching view of the relationship between BT's risk, its initial capital expenditures in creating the 21CN, its lower overall unit costs once 21CN is fully operational and once legacy SMP offerings are no longer required in

**24** *Consultants' Report*, page 28 and page 36.

parallel, and the desired course for regulated rates for narrowband voice interconnection (and presumably to other regulated prices) as a result:

> … IP voice interconnect charges would need to start above the costs of a hypothetical stand alone NGN, because to do otherwise would create an arbitrage opportunity where (for example) migration costs would not get recovered. However, these IP voice products could still be priced below C7/TDM narrowband interconnect products to the extent they cost less to provide than C7/TDM interconnect products. Finally, at a point in the future, when all traffic is via IP voice interconnect, and all migration / PSTN costs had been recovered, IP interconnect pricing would end up reflecting the costs of the NGN allowing an appropriate rate of return.[25]

Figure 5: 'Holistic' approach to narrowband voice interconnect cost recovery (illustrative only)




Source: Ofcom, *Further Consultation*, Figure 5, page 13.

This intriguing diagram represents a fascinating thought model, but it also raises many questions that do not appear to be explicitly answered in the Ofcom documents.[26]

The upper line, "NCC. Based on theoretical PSTN-only network" is the expected trend for the Network Charge Control (NCC) for BT's existing wholesale interconnect (discussed at more length in Section 2.3.2.2). It declines over time because BT's

**25** *Further consultation*, section 3.10.
**26** Much of what follows comes from discussions with Ofcom.

efficiency is presumed to improve over time. It is implicitly assumed that the efficiency of a network that is part PSTN and part NGN will improve no less quickly than BT's current PSTN network. In the event that the migration to NGN enables still greater efficiency gains, then BT reaps the benefit over the defined lifetime of these cost controls, which is 2005-2009 – the NCC level will not be revised other than in exceptional circumstances.[27]

The next line below, "IP voice interconnect charges", represents an as-yet-undefined NCC for a new wholesale SMP product enabling interconnection to narrowband voice services. It is presumably some form of IP interconnection. Given that this interconnect offering is not yet defined, much less implemented, the level of these charges has not yet been set; however, the general notion is that they should be less than those of traditional voice interconnect charges, but still sufficiently in excess of incremental cost to enable BT to recover the cost of migration from the PSTN to the NGN.

In fact, setting the price for IP-based interconnect to narrowband voice involves a number of complex decisions. First, the narrowband voice service at the *retail* level would appear to be in a single market, whether delivered over PSTN or NGN. Whether the same is true for interconnect to narrowband voice at the *wholesale* level is less clear.

More significantly, the question of NCC levels during a period of coexistence between traditional interconnect and new IP-based interconnect is complex. To the extent that network costs are lower, the interconnect price for IP-based interconnect should be lower; however, maintaining different interconnection prices for the same service will drive customers of the wholesale service to the new IP-based mode of interconnection. Should this be viewed as a positive industrial policy, or as an invitation to regulatory arbitrage? It is too soon to say.

#### 2.3.2.1 BT's cost of capital

Meanwhile, Ofcom has engaged in a lengthy series of proceedings to determine BT's appropriate cost of capital for purposes of regulatory rate-setting. These proceedings are based on BT's network as it exists today; however, they establish a framework that is intended to be carried forward into the world of the 21CN. We return to this theme at the end of this section.

---

**27** It is generally recognized that regulators should refrain from frequent or arbitrary reductions in regulated rates, due to the risk of reducing the incentive for operators to invest in efficiency improvements . Cf. Laffont and Tirole, *Competition in Telecommunications*, 2001.

The Weighted Average Cost of Capital (WACC) is a key driver in rate setting. It reflects the cost of equity, the cost of debt, and the company's gearing (a measure of the company's ratio between debt and equity).

Ofcom uses the Capital Asset Pricing Mechanism ("CAPM") to reflect risk and its impact on the returns that shareholders should expect.**28** CAPM has some known deficiencies, but it is widely used and theoretically well grounded. In CAPM, the cost of equity capital is rolled up from three components: (1) the risk free rate; (2) the expected market equity risk premium; and (3) the value of beta for the company in question. The CAPM makes no allowance for unsystematic (company specific) risks – investors are assumed to hold diversified portfolios such that these risks net out overall to zero in the expected case.

Ofcom's consultation related to the second and third of these factors: the equity risk premium (ERP) and the beta, respectively. The first factor, the Risk Free Rate (RFR), is simply the return that an investor would expect on a risk free investment.

The ERP is a stock-market factor, rather than being company specific. It reflects the degree to which investors expect a higher return for putting money into equity instruments (stocks) than into risk free investments. Ofcom determined that values between 4% and 5% were reasonable, and decided to use a value of 4.5%. This value is 0.5% lower than their previous value of 5%.

In calculating a beta for BT, Ofcom decided for the first time to use different values of beta for different major BT activities. Thus, while they computed an overall group beta of 1.1 for BT, they have chosen to "disaggregate BT's group beta of 1.1 into two components which broadly relate to BT's copper access network business with an equity beta of 0.9 and the rest of BT (including retail calls, broadband, and leased lines) with an equity beta of 1.23."**29** In other words, there is less business risk associated with BT's copper access network business than with the services that ride on top of that network.

Based on these disaggregated betas, Ofcom then estimates the weighted average cost of capital ("WACC") for BT's two component parts on a pre tax nominal basis to be 10.0% for the copper access network business, and 11.4% for the rest of BT.**30**

Ofcom performed this analysis on BT's current network, not on the 21CN; however, in their initial consultation, they had implied that risks might be slightly higher for next generation core networks, and significantly higher for next generation access networks, than for BT's current network, and proposed to address these differences through a modeling mechanism known as Real Options. In the Final Statement, they expressed a

---

**28** This discussion of CAPM is a synopsis of the well written discussion in section 3 of the *Final Statement*.
**29** *Final Statement*, section 1.22.
**30** *Final Statement*, section 1.23.

willingness to consider comments from stakeholders going forward. "Should the theoretical case for real options be demonstrated, Ofcom would then determine in consultation with stakeholders whether and how best to put this into practise. Ofcom considers that Next Generation Access may prove to be such a case …"[31]

#### 2.3.2.2 Narrowband voice interconnect

In the *Review of BT's network charge Controls*, Ofcom carried the existing structure of regulated prices for narrowband services (e.g. voice) over with only minor changes. The prices for these services are generally associated with caps tied to the *Retail Price Index (RPI*). The current structure, and the proposed structure for the four years beginning on October 1, 2005 appear below:

Table 1: BT's Network Charge Controls for four years from 1 October 2005

| **Service** | **Current controls 2001-5** | **Proposed controls 2005-9** |
|---|---|---|
| Call termination | RPI – 10% | RPI – 5% |
| Call origination | RPI – 10% | RPI – 3.75% |
| Single transit | RPI – 13% for combined basket | RPI – 11.5% |
| Local-tandem conveyance | | Safeguard cap of RPI – 0% |
| Interconnection circuits (ISB) | RPI – 8.25% for combined basket; RPI + 0% sub-caps for each of ISB & PPP | RPI – 5.25% |
| Product management, policy and planning (PPP) | | RPI + 0.75% |
| DLE FRIACO | RPI – 7.5% | RPI – 8% |
| Single Tandem FRIACO | RPI – 8.75% | RPI – 8.5% |
| Inter-tandem conveyance and Inter-tandem transit | Safeguard cap of RPI – 0% | No control as no SMP |

Source: Ofcom, *Review of BT's network charge controls: Explanatory Statement and Notification of decisions on BT's SMP status and charge controls in narrowband wholesale markets*, page 4.

Ofcom clearly views these 2005 – 2009 arrangements as an interim step, one that may rapidly be rendered irrelevant as Ofcom proceeds to implement the 21CN. The

**31** *Final Statement*, section 1.30.

document establishes a few general principles, but carefully refrains (as previously noted) from detailed analysis of a 21CN environment that does not yet exist. "…the new NCCs may be the last ones to regulate BT's … current set of narrowband services using the same broad approach that has applied since 1997. As BT changes to its new 21CN network, new interconnect products will be introduced, and Ofcom will have to consider the impact in terms of how markets are defined and how BT's wholesale services should be regulated."[32]

> … The NCCs described in this document will apply to a period of transition including the migration of BT from its current public switched telephone network ("PSTN") to its proposed 21st Century Network ("21CN"). Ofcom's analysis is therefore designed to meet these considerations on a forward looking basis. … In order to achieve this, Ofcom has adopted a technology neutral model to determine the average unit costs of narrowband PSTN services over the period to 2009. This is a way to cope with the uncertain speed of traffic migration to the 21CN, and to incentivise efficient migration of that traffic. It also has the effect of using hypothetical levels of PSTN capital expenditure during a period when it is expected that BT will move from PSTN to internet Protocol (IP) investment as part of its proposed 21CN deployment. Ofcom has assumed within this hypothetical model that BT will continue to improve its PSTN efficiency levels in line with historical experience and international benchmarks, and has set an achievable efficiency target for BT that is at the high end of the range on which Ofcom consulted. … As such, Ofcom has attempted neither to forecast actual efficiency gains that BT might reap from its 21CN deployment nor to take into account BT's forecast parallel running costs of running down its PSTN capability while migrating to 21CN. Ofcom will consider how to take account of BT's 21CN efficiency if and when Ofcom determines a price for 21CN interconnect services, and in any future NCC [network charging arrangements] in the period from 2009 onwards. …"[33]

This must be viewed as embodying an appropriate level of regulatory restraint. At the same time, it means that there is little more to be gained from studying these regulatory findings, inasmuch as the hard decisions have all been deferred until the situation is more mature.

---

**32** Section 1.6.
**33** Ibid., sections 1.12, 1.13 and 1.14.

## 2.4 Other developments

In parallel with these BT-oriented developments, alternative operators of all varieties continue to evolve their offerings, their networks, and their interconnection arrangements.

One particularly interesting recent development is the XConnect Alliance, a consortium of VoIP service providers who are committed to settlement-free interconnection.[34] A range of VoIP service providers (generally smaller firms) participate, including Telio, VozTelecom, Gossiptel, SipMedia, Gradwell, Musimi, Yak, Broadband Phone, TelAppliant, VoicePulse Inc., IOL, blueface, OSI, DataPro, ANEW Broadband, and IPness.

This reflects an important trend that cuts across much of the dialogue between BT and Ofcom. That dialogue presumes a traditional view of call termination arrangements, where the calling party's network must make a payment to the called party's network in recognition of the latter's cost of operation.

In the world of the Internet, however, there is no such presumption. Peering agreements are generally bilateral contractual arrangements, and frequently involve no payment in either direction. In the Internet, it is impractical to attribute cost causation to another network's customers. The presumption instead is that each network should recover its costs from its own subscribers.

The migration to NGN represents the evolution of the traditional PSTN telephony network in the direction of the Internet. In parallel with this, the world of the Internet continues to evolve and develop.

In the UK, both developments are simultaneously evident.

As these two worlds interconnect, the very different charging models on which they are premised will collide.[35] The results of that collision remain unclear.

---

**34** See http://www.xconnect.net/alliance.html.

**35** An exchange that took place at the recent "NGN and Emerging Markets" workshop (op. cit.) poses an interesting illustration of this tension. Dr. Frank Schmidt of T-Com contended that the network operator needed to be compensated for the use of its assets; Mr. Thilo Salmon of Sipgate countered that, in the case of an independent VoIP provider such as Sipgate, the network was merely moving raw data traffic, for which it was already being paid by its subscriber.

# 3 Developments in the US

In the United States, the evolution of traditional networks to IP is well along; however, that evolution takes a somewhat different course than that which is envisioned in Europe. Moreover, Americans do not think of this as a migration to the NGN – for the most part, NGN migration is thought to be an international issue, not a domestic one.

This is not just a semantic difference. Integrated networks carrying both voice and IP-based data over a common DWDM core have been common in the U.S. for many years. Service providers perceive no compelling business reason to evolve their data networks in a direction different from that of the familiar Internet. U.S. operators generally stopped buying traditional voice switches without VoIP capability years ago, but the integration of VoIP into existing networks has been gradual and incremental. In the U.S., the migration to NGN is merely a routine continuation of trends that have long been evident.

Interconnection of these IP-based networks is for the most part following “traditional” lines. IP interconnection based on peering is mature and stable, but does not provide for differentiated levels of Quality of Service in support of real-time bidirectional voice (and video). VoIP traffic is often carried on a single IP network, and then handed off as traditional PSTN traffic. This is unlikely to change any time soon (with the possible exception of VoIP handoff between cable TV providers)[36]. The factors inhibiting evolution have nothing to do with technology – the technology to hand IP traffic off at different levels of QoS have for the most part been available for a decade. Rather, economic incentives have been insufficient to drive deployment.

The United States government recognizes the need to spur the deployment of broadband, but generally refrains from measures to otherwise foster the evolution of IP-based networks (such as IPv6, DNS security, or inter-provider QoS). In terms of the manner in which these debates are framed in the U.S., to refer to a proposed initiative as “industrial policy” is to denigrate it.

## 3.1 Intercarrier compensation

Arrangements for intercarrier compensation for fixed and mobile telephony in the U.S. are extremely complex; nonetheless, the collective effect of this inelegant and rather Byzantine system has been to encourage some of the lowest mobile call termination rates in the world. Those low wholesale rates for mobile termination have led in turn to low retail rates, and to high utilization of mobile phones. In terms of economic welfare, this author would contend that the U.S. system is clearly superior to that of Europe.

---

**36** See http://www.lightreading.com/document.asp?doc_id=85991.

The arguments on these points have grown rather old and shop-worn, so we will not dwell needlessly on them here.[37] Our intent is to provide just enough background to enable the reader to make sense of U.S. attitudes toward compensation for traffic to and from the Internet, including VoIP traffic.

### 3.1.1 General mechanisms

First, we need to informally define a few terms:

- Local Exchange Carrier (LEC): a wired local exchange carrier.
- ILEC: an incumbent LEC.
- Regional Bell Operating Company (RBOC): an ILEC that was once part of the Bell system.
- LATA: a local calling area.
- intra-LATA call: a local telephone call.
- Inter-LATA call: a long distance telephone call.
- Inter-eXchange Carrier (IXC): a long distance carrier.
- Calling Party's Network Pays (CPNP): a system of compensation where network of the party that originates a telephone call must make a wholesale payment to the network of the party that receives (terminates) the call.

For local calls, there are two main rules that come into play. First, ILECs are generally restricted by regulation at the state level to termination rates that are cost-based. Second, companies can establish termination rates at any mutually acceptable level, as long as (1) those that are ILECs conform to the first rule, and (2) the rates are *symmetric*, i.e. the same in both directions. For purposes of these rules, mobile operators are treated exactly the same as CLECs (new entrant wired operators).

In practice, these rules have led to low termination rates for termination to or from ILECs (at a small fraction of a U.S. cent per minute), and in most cases to *zero* rates

---

**37** For a more detailed view of call termination in the U.S., see my earlier paper, "Call termination fees: the US in global perspective", presented at the 4th ZEW Conference on the Economics of Information and Communication Technologies, Mannheim, Germany, July 2004. Available at: ftp://ftp.zew.de/pub/zew-docs/div/IKT04/Paper_Marcus_Parallel_Session.pdf. See also S. C. Littlechild, "Mobile Termination Charges: Calling Party Pays versus Receiving Party Pays", to appear in *Telecommunications Policy 2006.*

(known as Bill and Keep) for most traffic exchanged among CLECs and mobile operators.

Why is this? Non-ILEC carriers can mutually agree to set the rates at any level. If traffic is roughly symmetric, and if rates must be equal, then the net flow of payments at the wholesale level will be negligible whether rates are high or low. Since carriers view these rates as part of their marginal cost of providing service (even though they net to zero), they will tend to set them low (usually zero so as to avoid the cost of accounting for them).[38]

Long distance carriers have historically been paid directly by retail consumers, and have made termination payments to any wired local exchange carriers on both the origination and the termination side. Through a bizarre asymmetry in the rules, mobile operators have never had a regulatory right to demand termination fees for long distance traffic. For RBOCs, these rates have generally been pegged to rates in the neighborhood of half a U.S. cent per minute. Rural ILECs may be permitted to charge significantly higher rates. A CLEC is not permitted to charge a higher rate than the ILEC for the geographic area in question (unless the CLEC wishes to demonstrate that its costs are higher than the ILEC – in practice, this is not done).

The combined effect of these rules, and of the commercial considerations that flow from them, is that call termination fees are often zero. Where they are not zero, they are usually very low. With the exception of a few rural ILECs and CLECs, carriers charge no more than 0.65 U.S. cents per minute.

These low wholesale rates have fostered low retail rates. Moreover, they have enabled flat monthly rates, which are increasingly common for wireless and for wired service.

Many Europeans have the mistaken impression that U.S. call termination arrangements flow from the regulatory imposition of costs at the retail level on the party receiving the call (Receiving Party Pays). In reality, no such regulation ever existed. U.S. call payment arrangements at the retail level are an economic consequence of the symmetry and parity obligations at the wholesale level.

Moreover, as the entire U.S. system migrates to flat rate plans, most consumers perceive their retail price per call minute as zero.[39]

The U.S. call termination system avoids a number of economic distortions that occur in most other countries. It does not generate artificial subsidies from fixed users to mobile – this has resulted in a slower maturation of the mobile market, but still an adequate

---

**38** Cf. Laffont and Tirole, *Competition in Telecommunications*, MIT Press, 2001.

**39** This perception is not altogether accurate. If they consume a few more minutes, their bill may not change. If, however, they consume a great many additional minutes, they may need to purchase a larger bucket of minutes.

development. Mobile penetration rates in the U.S. are currently at 65%, and growing at a rate of about 5% year over year.[40] The U.S. is thus within two or three years of falling within European norms.

The U.S. system has also solved another huge potential distortion: roaming charges. While roaming charges comprised 14% of U.S. mobile revenues in 1995[41], they represented just 5% of mobile revenues in 2002, and 4% in 2003[42].

The overall results are a call termination regime that generally requires regulatory intervention only for wired incumbents (a low level regulatory intervention), and that achieves some of the lowest retail prices in the world. Table 2 below, and Figure 6 which is based on it, show minutes of use in both directions and revenue per minute (which serves as a proxy for price per minute of use) for mobile users in a number of countries. The U.S. (and Canada, which has a generally similar system) are characterized by very low prices and extremely high minutes of use per month.

Table 2: Revenue per minute versus monthly minutes of use.[43]

| Country | Revenue per Minute ($) | Minutes of Use |
|---|---|---|
| USA | 0.08 | 630 |
| Hong Kong | 0.06 | 387 |
| Canada | 0.11 | 359 |
| South Korea | 0.10 | 316 |
| Singapore | 0.10 | 282 |
| Finland | 0.16 | 258 |
| France | 0.17 | 225 |
| Australia | 0.21 | 168 |
| Japan | 0.32 | 154 |
| UK | 0.22 | 151 |
| Spain | 0.27 | 135 |
| Italy | 0.26 | 120 |
| Germany | 0.35 | 76 |

Source: FCC, *Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services, 10th Report (10th CMRS Competition Report)*, July 2005, Table 10, based on Glen Campbell et al., *Global Wireless Matrix 4Q04*, Global Securities Research & Economics Group, Merrill Lynch, Apr. 13, 2005.

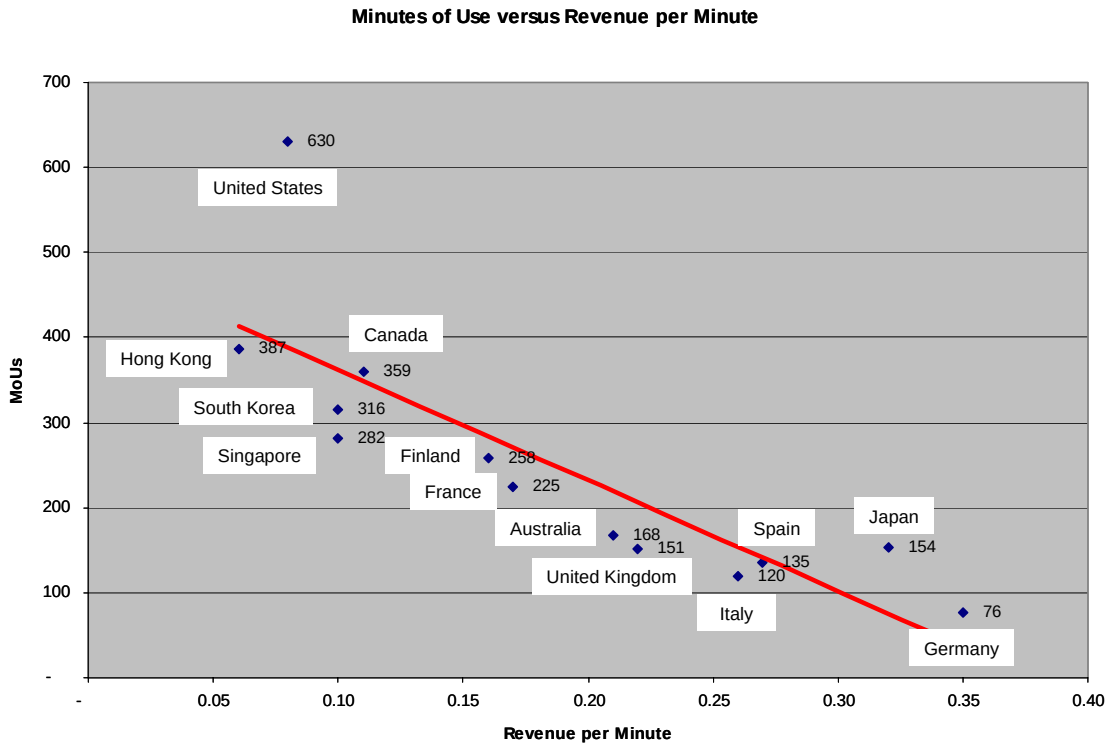
**40** FCC, *10th CMRS Comptetition Report*, July 2005.

**41** Cellular Telecommunications and Internet Association, *Semi-Annual Wireless Industry Survey* (see http://www.wow-com.com/industry/stats/surveys/).

**42** Ibid.

**43** FCC, *Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services, 10th Report (10th CMRS Competition Report)*, July 2005, Table 10, based on Glen Campbell et al., *Global Wireless Matrix 4Q04*, Global Securities Research & Economics Group, Merrill Lynch, Apr. 13, 2005.

Figure 6: Minutes of use versus revenue per minute




Source: The data derive from FCC, *Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services, 10th Report (10th CMRS Competition Report)*, July 2005, Table 10, based on Glen Campbell et al., *Global Wireless Matrix 4Q04*, Global Securities Research & Economics Group, Merrill Lynch, Apr. 13, 2005.

For a variety of reasons, the slope of the regression line should not be taken to represent demand elasticity. These are not the same consumers, and they do not have the realistic option of substituting one of these services with another. Nonetheless, it is fair to say that the data are highly *suggestive* of elastic demand.

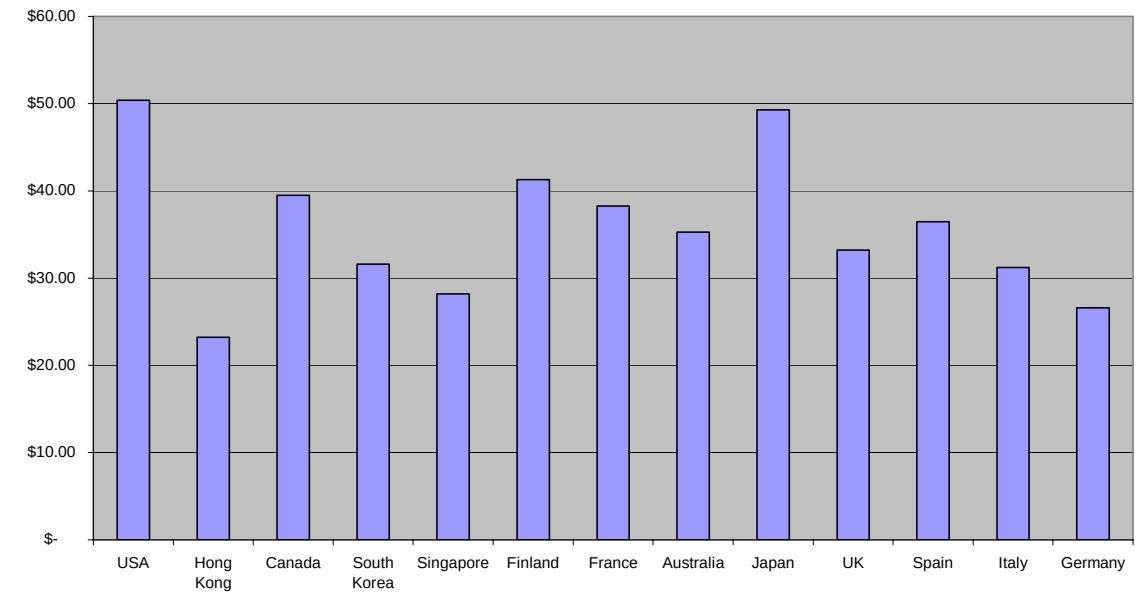
These data illustrate that consumption of mobile minutes per mobile user is *more than eight times higher* in the United States than in Germany.[44] The low rate of consumption in Germany (apparently reflecting depressed rates of outbound calls due to the relatively high price per minute) arguably reflects a deadweight social loss. This study did not explicitly consider the German market, but these data suggest that German consumers would probably be better off if Germany were further to the left and higher

**44** 630 minutes of user per month versus 76, for a ratio of 8.29:1.

on this graph (with lower revenues per minute, and a correspondingly higher number of minutes consumed per month.)

This loss of consumer welfare is not necessarily offset by provider welfare. Figure 7 shows that ARPU is notably higher in the U.S., with low revenue per minute, than it is in Germany, with high revenue per minute.[45]

Figure 7: Average monthly revenue per unit (ARPU)




Source: The data derive from FCC, *Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services, 10th Report (10th CMRS Competition Report)*, July 2005, Table 10, based on Glen Campbell et al., *Global Wireless Matrix 4Q04*, Global Securities Research & Economics Group, Merrill Lynch, Apr. 13, 2005.

The FCC has historically favored low or zero termination rates, reasoning that low termination rates at the wholesale level would encourage low retail prices. These same low termination rates also made possible the evolution to mobile plans with large bundles of minutes, and to fixed line and VoIP plans with unlimited minutes.

Patrick de Graba presented the idea in this way in a widely read FCC white paper[46]:

**45** Computed from the same data. Note, however, that this reflects gross revenues, not profitability. An analysis of the actual impact on provider profitability is well beyond the scope of this paper.

> One source of inefficiency is that existing termination charges create an "artificial" per-minute cost structure for carriers that will tend to result in inefficient per-minute retail prices. In unregulated, competitive markets, such as the markets for [mobile telephony] services and Internet access services, retail pricing is moving away from per-minute charges and towards flat charges or two-part tariffs that guarantee a certain number of free minutes. This suggests that few costs are incurred on a per-minute basis, and that flat-rated pricing will lead to more efficient usage of the network. The existing reciprocal compensation scheme, which requires the calling party's network to pay usage sensitive termination charges to the called party's network, imposes an "artificial" per-minute cost structure on carriers which, if retail rates are unregulated, will likely be passed through to customers in the form of per-minute retail rates. Such usage sensitive rates thus would likely reduce usage of the network below efficient levels.

### 3.1.2 The Enhanced Service Provider Exemption

The United States has long exempted telephone calls from the consumer to the Internet Service Provider from long distance access charges that would otherwise be due for calls outside of the metropolitan area. Under the Enhanced Service Provider (ESP) Exemption, they are treated as jurisdictionally interstate (in order to assert exclusive FCC jurisdiction, and thus to prevent the states from imposing charges), but are treated for charging purposes as local calls.[47]

Since local calls in most parts of the United States can be made subject to unmetered plans, this had the effect of eliminating per-minute charges for most calls to connect to the Internet.

This had a number of complex but fairly predictable results. On the one hand, the absence of per-minute charges encouraged use of the Internet, and fostered rapid growth.[48] At the same time, the regulatory asymmetry between these calls and other calls has created many opportunities for arbitrage and mischief. The FCC has managed to "patch" the most egregious problems, but the irrationalities and asymmetries in the call termination system are likely to continue to cause problems.

---

**46** See Federal Communications Commission (FCC) Office of Strategic Planning and Policy Analysis (OSP) Working Paper 33: Patrick DeGraba, "Bill and Keep at the Central Office As the Efficient Interconnection Regime", December 2000, at 95, available at http://www.fcc.gov/osp/workingp.html.

**47** The basis on which the call is billed is the distance from the consumer to the point of connection to the ISP (which is local in most but not all cases), not to the end destination of the traffic.

**48** In this sense, it is somewhat akin to systems like FRIACO in the UK. At the same time, the presence of cheap unmetered dial-up access to the Internet may have paradoxically served to slow adoption of broadband in the US – consumers had less incentive to upgrade.

A further complicating factor is that the United States funds universal service partly through a levy imposed on all providers of telecommunication services, and partly through implicit subsidies in the structure of call termination rates. The migration of call traffic to the Internet is thus one of several factors that is causing a shortfall in the funding of universal service.

### 3.1.3 U.S. interest in migration to Bill and Keep

A number of U.S. experts (many of them with connections to the FCC) have argued that the way to achieve consistency between Internet wholesale compensation models and those of the PSTN, in order to avoid arbitrage in the future converged world, is to evolve the PSTN models in the direction of those used in the Internet.[49] For the United States, this is not such a radical notion – the compensation arrangements among mobile operators and CLECs are for the most part already there.

The arguments for Bill and Keep rest on three primary pillars[50]:

1. Rejection of the notion that the party that places a call should be viewed as the sole cost-causer;

2. Avoidance of the classic terminating monopoly problem, without the need to impose regulatory caps on the termination rates of all service providers; and

3. The need to establish a uniform system of call termination in order to avoid damaging regulatory arbitrage.

In the PSTN world, the party originating the call generally pays for the call, while the party receiving the call typically pays nothing. In the Internet, by contrast, each end customer generally pays for all connectivity to the customer premises – indeed, it would be difficult if not impossible to allocate responsibility for individual traffic flows.[51] As De Graba puts it, "… both parties to a call – i.e., the calling party and the called party –

---

**49** See De Graba, op. cit, and also FCC Office of Strategic Planning and Policy Analysis (OSP) Working Paper 34: Jay M. Atkinson and Christopher C. Barnekov, "A Competitively Neutral Approach to Network Interconnection", December 2000. Both papers are available at http://www.fcc.gov/osp/workingp.html. My own published work is also in this tradition.

**50** See De Graba, op. cit., at 4.

**51** See also Jean-Jacques Laffont, J. Scott Marcus, Patrick Rey, and Jean Tirole, IDE-I, Toulouse, "Internet interconnection and the off-net-cost pricing principle", *RAND Journal of Economics*, Vol. 34, No. 2, Summer 2003. An earlier version of the paper is available at http://www.idei.asso.fr/Commun/Articles/Rey/internet.pdf. "Finally, let us compare [these results] with the results in Laffont, Rey, and Tirole (1998a) and Armstrong (1998) for interconnection of telephone networks. A key difference with this telecommunications literature is that in the latter there is a missing price: receivers do not pay for receiving calls… In sum, the missing payment affects the backbones' perceived costs, and it reallocates costs between origination and reception."

generally benefit from a call, and therefore should share the cost of the call.[52] By requiring interconnecting networks to recover most, if not all, of the cost of the call from their own customers, [Bill and Keep] provides an efficient means by which the parties to a call can share the total cost of a call."[53]

De Graba goes on to argue "… that competition operates more effectively when carriers recover their costs from their own end users, who can choose among competing carriers, rather than from interconnecting networks for whom the terminating carrier is a de facto monopolist. [Bill and Keep] takes advantage of the forces of competition, where they exist, by requiring a carrier to recover all of its local access costs from its end users."[54]

The last argument may be the most telling. The current PSTN call termination arrangements on both sides of the Atlantic reflect significant economic distortions. The Internet represents an absolutely classic case of high initial costs and low marginal costs. Moreover, the Internet makes regulatory bypass trivially simple. Internet-based bypass will over time defeat any attempt to maintain usage-based prices substantially in excess of marginal cost, which is practically nil.[55]

The FCC has been trying for years to migrate to a uniform call termination system premised on overall notions of Bill and Keep.[56] In the highly politicized regulatory environment of the United States, they have been unable to make headway against the determined opposition of those carriers whose financial interests would be impacted by such a migration. The large fixed incumbent operators (called RBOCs) have for the most part been reasonably supportive of a migration to Bill and Keep; small rural fixed operators, whose termination charges tend to be much higher, have been the main opponents.

A significant implication for Germany is that this is an issue where it is difficult to develop a consensus – inevitably, any significant change is likely to negatively impact some operators, who will therefore oppose the change vigorously.

The FCC attempted re-launch this proceeding by issuing a *Further Notice of Proposed Rulemaking (FNPRM)* on March 3, 2005.[57] The statements that the FCC commissioners issued at the time make it clear that the issue is becoming increasingly

---

**52** See also Doh-Shin Jeon, Jean-Jacques Laffont, and Jean Tirole, "On the receiver pays principle", *RAND Journal of Economics*, 2004. They explore the inherent mirror-image relationship between calling and called party, and find that there is no qualitative difference, as "it takes two to tango."

**53** De Graba, op. cit., at 4.

**54** Ibid.

**55** The FCC made exactly this point in the Universal Service Report (the "Stevens Report") in 1998, when they noted that "… IP telephony serves the public interest by placing significant downward pressure on international settlement rates and consumer prices."

**56** FCC, *In the Matter of developing a Unified Intercarrier Compensation Regime*, CC Docket 01-92, released April 27, 2001.

**57** Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-33A1.pdf.

urgent; nonetheless, there is no obvious progress to date on the FNPRM. To a certain extent, the FNPRM might be viewed as a procedural rather than a substantive matter – had the FCC attempted to issue a ruling based on the old proceeding without providing a new opportunity for stakeholders to comment, they would have risked being overturned by the courts on purely procedural grounds.

## 3.2 Technical and economic background on IP Quality of Service (QoS)

The technical capability to upgrade the peering connections among Internet backbones in order to provide enhanced Quality of Service (QoS) has existed for a decade; nonetheless, there is negligible deployment of these capabilities between independently managed service providers.

In order for the reader to make sense of these industry developments – or of the lack of developments – it is necessary to introduce some technical and economic background, and to so from a U.S. perspective. This section of the report seeks to do so.

### 3.2.1 Peering and transit

The best explanations of the primary forms of Internet interconnection (*peering* and *transit*) appear in a publication of the Network Reliability and Interoperability Council (NRIC), an industry advisory panel to the FCC:

> *Peering* is an agreement between ISPs to carry traffic for each other and for their respective customers. Peering does not include the obligation to carry traffic to third parties. Peering is usually a bilateral business and technical arrangement, where two providers agree to accept traffic from one another, and from one another's customers (and thus from their customers' customers). …
>
> *Transit* is an agreement where an ISP agrees to carry traffic on behalf of another ISP or end user. In most cases transit will include an obligation to carry traffic to third parties. Transit is usually a bilateral business and technical arrangement, where one provider (the transit provider) agrees to carry traffic to third parties on behalf of another provider or an end user (the customer). In most cases, the transit provider carries traffic to and from its other customers, and to and from every destination on the Internet, as part of the transit arrangement. In a transit agreement, the ISP often also provides ancillary services, such as Service Level Agreements, installation support, local telecom provisioning, and Network Operations Center (NOC) support.

> Peering thus offers a provider access only to a single provider's customers. Transit, by contrast, usually provides access at a predictable price to the entire Internet.
>
> Historically, peering has often been done on a bill-and-keep basis, without cash payments. Peering where there is no explicit exchange of money between parties, and where each party supports part of the cost of the interconnect, … is typically used where both parties perceive a roughly equal exchange of value. Peering therefore is fundamentally a barter relationship.[58]

There has been a tendency in the literature to assume that all peering is "free". This is simply not the case. When the author was in charge of peering policy for GTE Internetworking (at the time one of the five largest Internet backbones in the world), about 10% of our peering relationships involved payment. These payments had nothing to do with the relative size of the participants; rather, they were a reflection of traffic imbalance. For Internet backbones interconnected at multiple points by means of shortest exit routing, the traffic *received* from another network must on the average be carried further, and must therefore cost more, than the traffic *sent* to the other network.[59]

### 3.2.2 Efforts to provide end-to-end Quality of Service (QoS)

By the early Nineties, it had already become obvious to the engineering community that real-time bidirectional voice and video communication could potentially benefit from delivery guarantees on delay. This led to a series of standards efforts – first, the *RSVP*-based Integrated Services Architecture, and then to *Differentiated Services (DiffServ)*.

RSVP provided a comprehensive end-to-end QoS management architecture. Over time, it came to be viewed as hopelessly complex,[60] and was effectively abandoned in favor of DiffServ. DiffServ provides a simple means of specifying, on a hop-by-hop basis, the desired performance characteristics – it is then up to the network to meet those requirements as well as it can.

DiffServ should thus be viewed as a *signaling* mechanism. Technically, it is trivial. The implementation of QoS *within* an IP-based network, with or without DiffServ, has been straightforward with or without DiffServ for at least a decade. Implementation of QoS

---

**58** Report of the NRIC V Interoperability Focus Group, "Service Provider Interconnection for Internet Protocol Best Effort Service", page 7, available at http://www.nric.org/fg/fg4/ISP_Interconnection.doc.

**59** Ibid., pages 4-6. See also Marcus, *Designing Wide Area Networks and Internetworks: A Practical Guide*, Addison Wesley, 1999, Chapter 14.

**60** This is not altogether true. My former firm, BBN, operated a commercial RSVP-based network for many years. It was a commercial failure, but not a technical failure.

*between or among* independently managed IP-based networks has never gotten off the ground. Given that the technology is fairly simple, the answers clearly lie in business and economic factors.

In the balance of this section of this report, we review a number of the business, technical and economic factors that have collectively slowed the adoption of differentiated QoS among providers.[61] Key considerations include:

- Today's routine best-efforts service works well enough under most circumstances most of the time; consequently, consumer willingness to pay a substantial premium for better service is low.
- The value of better-than-best-efforts service increases as more destinations are reachable using the service (a property known as a *network effect* or *network externality*); however, it is difficult to get past the initial adoption hump.
- Achieving widespread deployment tends to require agreement of many parties. The costs of obtaining those agreements, and the time needed to do so, reduce the effective economic return to the network operator that must invest in making the service available.
- For all of these reasons, network operators in the U.S. have found it difficult to justify investment in differentiated QoS.

It is also worth noting that, in the context of the United States, attempts to implement differentiated QoS between providers have generally been in connection with *peering*. It is conceivable that a large German operator might find it easier to implement differentiated QoS services in connection with a *transit* service, where the financial arrangements between transit provider and transit customer are more clear-cut.

#### 3.2.2.1 Application requirements

There is a temptation to assume that all voice and video traffic requires assured quality of service.

This is not exactly correct. The receiving application typically implements a *jitter buffer* that can be used to smooth the variability in end to end delay. For streaming (one way) audio or video, most users will tolerate a delay of a few seconds when the application starts up. After that, a jitter buffer can typically deal with a considerable amount of variable delay.

---

**61** This argument is presented in greater depth in Marcus, "Evolving Core Capabilities of the Internet", *Journal on Telecommunications and High Technology Law*, 2004

Real time bidirectional voice and video pose a much greater problem. It has long been understood that, where the end to end delay exceeds about 150 to 200 milliseconds, users will “collide”. They will both start speaking at roughly the same time, because neither hears initially that the other is speaking. Those of us who remember international telephone calls routed over satellites are familiar with the phenomenon. This imposes a practical ceiling on the delay that the jitter buffer can allow.

This delay in turn imposes limits on both the mean and the standard deviation of delay for the traffic. In an IP-based network, the traffic is composed of individual packets. The delay for these packets can be viewed as comprising a fixed component (based primarily on the speed of signal propagation along the path from send to receiver, and thus dependent primarily on the distance along the path, and also on the deterministic delay to “clock” the packet onto each outbound data transmission link) and a variable component (based on queuing delays in each router through which the packet must pass, especially those associated with gaining access to the outbound transmission link). For a given traffic flow, the unidirectional delay can thus be viewed as a probability distribution with a mean and a standard deviation.

The ability to achieve a round trip delay of not more than 150 milliseconds depends on both the mean and the standard deviation of delay. It is a classic statistical *confidence interval* problem – it is necessary that the “tail” of the distribution in excess of about 150 milliseconds be suitably small. Note that an occasional outlier is generally permissible – as an example, the *codecs* (coder-decoders) used for Voice over IP (VoIP) services typically interpolate over missing data, and the human ear does a surprisingly good job in compensating for very short data losses. Human speech presumably incorporates a great deal of redundant information that can be used to fill in the gaps.

#### 3.2.2.2 Network performance in relation to application needs

We consider both fixed and variable delay. Fixed delay is comprised primarily of propagation delay and clocking delay.

We often forget that the speed of light *is* a meaningful constraint. In vacuum, light travels about 300 Km in a millisecond. Signal is not quite as fast when propagating through wires or fiber; moreover, the fiber runs do not proceed in a geometric straight line. For international calls, propagation delay can consume a significant fraction of the 150 millisecond budget.

Clocking delay is a function of the speed of the transmission link. Over a dial-up connection to the Internet, clocking delay poses a serious constraint. Over broadband media, it is much less of an issue. In the core of the Internet, the links are very fast indeed, so the deterministic clocking is correspondingly small.

Variable delay is best modeled and analyzed on a hop by hop basis. At each hop, it primarily reflects the queuing delay waiting to clock the traffic onto an outbound link. (Queuing delay for the processor of the router is also possible, but unless the processor is saturated it is generally small enough to ignore.) This variable delay can be analyzed using a branch of mathematics known as *queuing theory* – the science of waiting lines.[62]

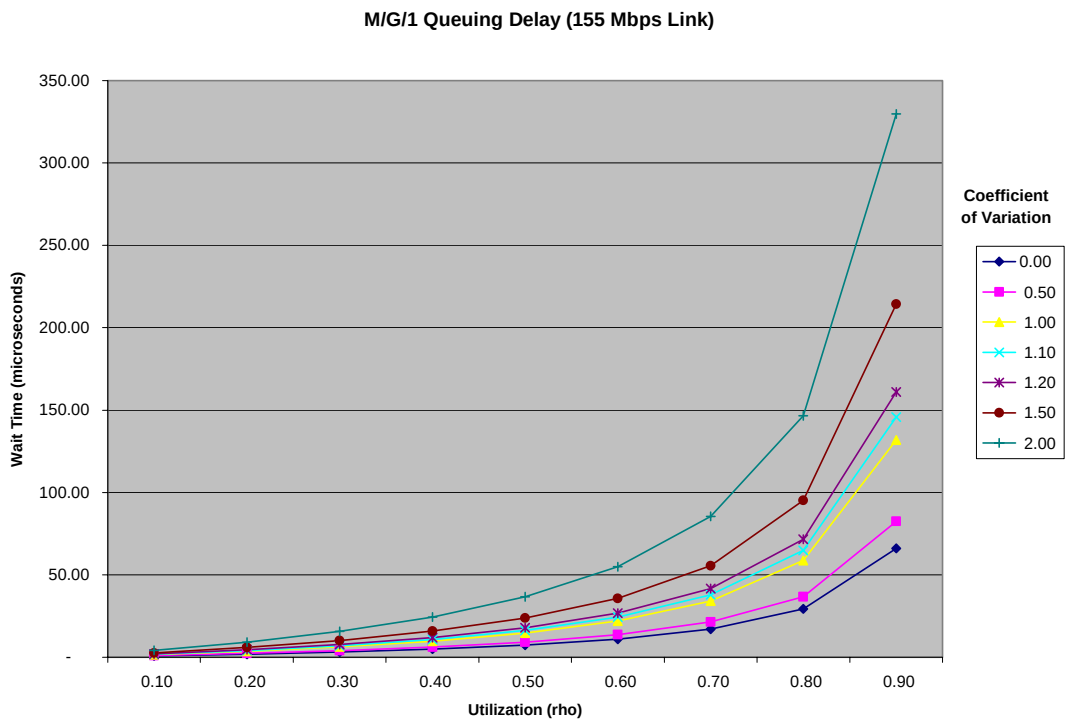
Queuing theory tells us that average variable delay reflects three things:

- The average service time (in this case, the deterministic clocking delay);
- The load on the server, which we can think of as the percent of time that it is busy; and
- The variability of the service time, expressed as a coefficient of variation (the standard deviation divided by the mean).

What queuing theory tells us about variable delay in the core of the large IP-based networks is that, *in a properly designed network and under normal operating conditions,* it plays only a very minor role. Figure 8 below depicts the average packet wait time for a 155 Mbps data link, which is the *slowest* link that one would expect to find in the core of a modern Internet backbone.

**62** For an introduction to the use of queuing theory in this context, see Chapter 16 of my textbook, *Designing Wide Area Networks*, op. cit.

Figure 8: Packet wait time on a 155 Mbps link




Source: Own computations

Among the family of curves shown, the one corresponding to a coefficient of variation of 1.20 is the one the accords most closely with observational experience around 2001, the most recent date on which this author had access to industry statistics.[63]

The computed average wait time per hop, even at a utilization of 90%, is about 150 *microseconds*. Note that this is *three orders of magnitude* less than our delay budget of 150 *milliseconds*. Beyond this, consider that many backbone links today are one or two orders of magnitude faster than 155 Mbps, with predicted delays correspondingly smaller.

This is not to say that delay could never be a problem. The same queuing theory analysis tells us that, as utilization approaches 100%, predicted mean wait time increases with no upper bound. But no network should be *designed* to operate routinely at those levels. Saturation will occur either as a result of (1) poor planning or forecasting

**63** The graph was computed using the Pollaczek-Khinchine formula for an M/G/1 queuing model. A mean packet length of 284 octets is assumed, consistent with observational experience around 2001.

on the part of the designer, or (2) substantial failures elsewhere in the network that necessitate re-routing of traffic.

#### 3.2.2.3 Commercial implications

The analysis in the preceding section tells us a great deal about the delay in deployment of QoS capabilities among Internet providers.

DiffServ-based QoS capabilities cannot speed up a network; they can only prevent it from slowing down (for certain packets) under load. They generally determine (1) which queued packets are served first, and (2) which queued packets are discarded when there is insufficient room to store them.

Under most circumstances, these effects will be too small for the end user to perceive.

It should come as no surprise that end users are unwilling to play a large surcharge for a performance improvement that is not visible to them.[64]

This is not to say that there is no commercial opportunity for inter-provider QoS; rather, it argues that the opportunities will not necessarily be found in the core of the network, which is the place where most people tend to look for them.[65] Instead, QoS will tend to be commercially interesting:

- For slower circuits at the edge of the network;
- For shared circuits to the end user (e.g. cable modem services);
- When one or more circuits are saturated;
- When one or more components have failed;
- When a force majeure incident has occurred; and especially
- Where more than one of these factors are present.

#### 3.2.2.4 An additional consideration: network externalities

A related concern has to do with the economics of *network externalities*. QoS is typical of capabilities that take on value only as more people adopt them.

---

**64** This was, of course, the key root problem in BBN's inability to successfully commercialize its RSVP-based commercial QoS-capable network.

**65** In a classic joke, a child looks for a lost coin under a lamp post, not because he lost it there, but rather because that is where the light is best.

The economist Jeffrey H. Rohlfs has written extensively on the subject, noting that many high technology services encounter difficulty in achieving sufficient penetration to get past an initial adoption hump.[66] Different successful offerings have met this challenge in different ways.

Certain Internet capabilities have deployed effortlessly – for example, the worldwide web. Conversely, others have tended to stall for reasons not necessarily related to technology, notably including IP version 6 (IPv6), DNS security (DNSSEC), and multicast. A common characteristic among the stalled capabilities is that, rather than being end to end features independent of the network, the stalled capabilities require concerted action and concerted change to the core of the network. Consequently, high transaction costs hinder initial deployment, and thwart attempts to reach critical mass and thereby to get beyond the initial adoption hump.[67]

Regrettably, inter-provider QoS seems to clearly fit the profile of the stalled capabilities.

## 3.3 Operational Support System (OSS) requirements for QoS

In this section of the report, we consider briefly the requirements that QoS would place on Operational Support Systems, and explain some of the approaches that were considered to upgrade the system.

One might assume that the lack of progress in the U.S. implies a lack of interest. This was not in fact the case. The various discussions that have taken place over the years shed a great deal of light on the challenges that any such system would face. Most of these challenges are best understand in terms of their systemic implications in terms of OSS implementation.

In order to understand the impact on OSS, it is necessary to begin by considering the *business* requirements and the associated *technical* requirements.

Today, there is less to be said about OSS *implementation* in support of accounting for NGN. In North America, charge accounting systems for peering are not deployed to any great extent because most peering is on a Bill and Keep basis (not explicitly charged for), while transit is charged based on the maximum carrying capacity of the “pipe”. In neither case is sophisticated usage-based charging required.[68] In Europe, the migration to NGN may drive much greater interest in accounting systems, but that migration is not yet advanced enough to have done so.

---

**66** Jeffrey H. Rohlfs, *Bandwagon Effects In High-Technology Industries* 3 (2001).

**67** I make this case at length in “Evolving Core Capabilities of the Internet”, *Journal on Telecommunications and High Technology Law*, 2004.

**68** There are solutions to verify compliance with Service Level Agreements (SLAs). See, for instance, http://www.brixnet.com/solutions/enterprise.shtml.

### 3.3.1 Intercarrier compensation charges

In particular, to understand accounting system requirements, it is necessary to begin by asking what it is that would be charged for.

If one Internet Service Provider (ISP)[69] is to expect another to honor its request to carry some particular stream of data at some preferred of QoS, it is safe to assume that the latter would expect some form of financial compensation from the former.

As a practical matter, this will almost certainly be associated with the volume of traffic in question. It is likely to be linked to the quality of service requested, but not explicitly to the application. The reasons for this are many:

- In a layered world, the ISP is not necessarily the application service provider. (Consider, for example, an independent VoIP provider such as Skype that operates transparently over any ISP transport facilities.)

- Application Layer characteristics will be visible to application service provider, but under layering they should not be visible to the underlying ISP.

- For an ISP, it is trivially simple to measure traffic across a link, and also to characterize by simple measures (port number, QoS requested). More sophisticated measures – developing an overall traffic matrix, for example – often involve large volumes of data, and significant complexity and expense.[70]

- If ISPs attempted to assess a surcharge significantly in excess of cost on application service providers, the latter would likely respond by encrypting their traffic or otherwise attempting to restrict visibility into the application.[71]

- Many of the traditional metrics used for billing – minutes of use, for example – are largely unrelated to cost causation in an IP-based network. If users and vendors are able to bypass non-cost-based charges, they will do so.

All of this suggests that intercarrier compensation is likely instead to be based on measures that the ISPs can directly measure, that are not easily forged, and that correlate with cost causation. This would appear to imply that the level of compensation

---

**69** In the context of this discussion, an NGN operator carrying IP-based traffic should be viewed as being an Internet Service Provider (ISP).

**70** More sophisticated measurements are possible, but they are not likely to be simple or inexpensive. See, for instance, www.narus.com.

**71** We assume that wholesale intercarrier compensation payments would most likely be reflected in prices at the retail level.

would likely reflect the volume of traffic, and that traffic associated with different levels of quality of service would be associated with different access charge rates.[72]

The ISPs could, of course, mutually agree to set any or all of these rates to zero. Where traffic is balanced and unit costs are similar, ISPs may tend to prefer a Bill and Keep system. Doing so avoids accounting complexity, and facilitates the use of flat rate pricing to the end user (because it avoids the risk of adverse selection).

### 3.3.2 Intercarrier compensation accounting

At a functional level, billing systems for intercarrier compensation have to support the relevant charging model. Under the assumptions presented in the previous section, this means that accounting systems would have to capture measures of traffic volume, segregated by the requested QoS.

In reality, there is a great deal of hidden complexity in such systems, even where the underlying functions seem to be very simple. Both parties must ultimately be able to agree on the charges assessed. This implies the existence of:

- Clear detailed bills, with supporting detail available if necessary to substantiate them.
- Means of demonstrating the degree to which the ISP receiving the traffic complied with the request.
- Pre-arranged tools, staff and procedures for resolving disputes when the statistics do not match up.

#### 3.3.2.1 Basic accounting

The need for detailed bills is obvious. The primary requirement is knowing the amount of data that one ISP received from the other, categorized by the QoS requested.

Not every QoS request will have been honored. We assume *arguendo* that QoS signaling would be implemented by means of DiffServ technology. With DiffServ, traffic is classified *or reclassified* on entry to a network. It can be metered, marked, policed, and shaped.

---

**72** This was our modeling assumption in Laffont et. al., “Internet interconnection and the off-net-cost pricing principle”, op. cit. It led to a relatively simple system with no obvious contradictions. Note that it is trivial for accounting systems to capture aggregate data on the number of packets sent or received on an interface, and to categorize those results according to the various levels of requested QoS.

#### 3.3.2.2 Demonstrating compliance

Finally, the sending ISP has no intrinsic way of knowing that the receiving ISP's network is actually designed to deliver the mutually agreed performance. At the end of the day, what is important is not the amount of traffic accepted, but rather the performance with which that traffic was carried. How is this to be verified?

DiffServ specifies Per Hop Behaviors (PHBs); however, the per hop queue management is really of no interest *per se* to the requesting ISP. The requesting ISP typically wants to know that its own customer received an overall end to end delay consistent with its expectations.[73] These expectations would likely be expressed contractually in the form of a Service Level Agreement (SLA).[74] The receiving ISP might have attempted to reduce its costs by undercapacitizing its network, in which case the traffic might encounter greater-than-intended delay.

A number of tools have evolved over the years for measuring delay. Historically, most ISPs have used a relatively primitive tool known as PING to measure round trip delay within their networks, even though more sophisticated tools exist.[75]

This discussion of tools is in reality but the tip of a very large iceberg. The verification of end to end delay is in practice a devilishly complex problem, for a number of reasons, most of which have more to do with business concerns than with technology:

- The need for mutual verification of end to end delay implies that two ISPs that are competitors for the same customers must share sensitive information about the internal performance of their respective networks. Each will worry that proprietary advantages will inadvertently be exposed, or conversely that weaknesses will be revealed and publicized.

- No ISP will want a competitor to directly access the routers and servers that its own customers depend on for the competitor's measurement purposes. This implies in turn the need to deploy new gear at a large number of locations solely to meet commitments to a competitor. This might be possible if both providers see some shared competitive advantage emerging, but the experience has shown that these deployments receive low priority.

- There is considerable uncertainty as to the end to end performance between providers that can be safely committed.

---

**73** Consider, too, that there may be additional ISPs downstream from the peering ISPs. In general, an ISP cannot guarantee performance beyond the boundaries of its own network.

**74** It is not necessary to measure the performance of every packet, but it is necessary at a minimum to be able to perform spot checks.

**75** See, for example, the IPPM-based measuring network maintained by RIPE NCC, at http://www.ripe.net/projects/ttm/.

- ISPs are understandably reluctant to enter into contracts with direct competitors that involve potential financial penalties if either party fails to meet agreed service levels.

- To the extent that financial penalties are involved, there may be incentives for one or both parties to game or manipulate the system. It is unlikely that all of the unpalatable scenarios can be predicted in advance.

All of the foregoing notwithstanding, it is not inconceivable that progress could be made. The author has advocated the establishment of inter-provider measurement systems as a first step, with the thought that advisory rather than mandatory SLAs might follow. Financial penalties would be agreed only after the parties had developed sufficient experience with the quality and stability of the measurement framework.

#### 3.3.2.3 Dispute resolution

The need for dispute resolution may be less obvious, especially in the most basic case.

One might imagine that simply measuring the traffic flowing over a link is so trivial as to need no further discussion, but experience says otherwise. Data reporting intervals will not align perfectly. If quantiles (percentiles) are reflected in the billing, then sampling intervals between both parties must be agreed.[76] One might expect that if one ISP measures the traffic heading away from it on a given circuit, and another measures the traffic heading towards it on the same circuit, the measurements should agree, but even this most basic assumption will not always hold.

There is an old Dutch proverb: “Never go to sea with two compasses. Take one, or three.”

Disputes will inevitably arise.[77] How are they to be dealt with?

One possibility would be to empower some impartial third party to capture statistics for both ISPs, to generate settlement bills, and perhaps to act as an honest broker for dispute resolution. It is not clear who, if anyone, the ISPs might trust to do this; beyond that, there would still be the need for some kind of fully independent appeal.

As a penultimate line of defense, some form of mediation or arbitration would clearly be preferable to having all such cases go to court.

---

**76** This is a subtle consequence of the Central Limit Theorem. The more frequent the sampling interval, the “lumpier” the data distribution will seem to be.

**77** As previously noted, about 10% of my former employer’s conventional peering arrangements involved compensation. Dispute resolution chewed up extravagant amounts of time.

These challenges do not necessarily lend themselves to a BNetzA-driven solution; commercial arrangements on the part of service providers should suffice. Nonetheless, continued dialog between BNetzA and industry on the various issues related to interconnection in an NGN world is appropriate.

# 4 Implications for Germany

Evolution in the U.S. and the UK is following notably different paths. In this section, we consider what specific actions for Germany might logically follow from experience in the US and in the UK.

## 4.1 The UK

The evolution in the UK represents an exciting and refreshing change. The evolution of BT's network, while potentially dramatic in scale, is not novel in concept. Much more significant is the massive overhaul of the regulatory framework that is currently under way.

Ofcom has, as usual, prepared a comprehensive and thoughtful set of consultation documents. They are capitalizing on the 21CN to overhaul the regulatory environment, while carefully maintaining consistency with the parallel operation of the European regulatory framework for electronic communications. Their efforts to place primary reliance on enforceable instruments to ensure equivalence of input at the wholesale level, while withdrawing from the retail level as much as possible, would appear to be logical and promising. The intent is to focus regulation on those areas where enduring bottlenecks exist.

Ofcom has for the most part deferred the hard decisions to the future. This is perhaps appropriate – it may simply be too early to define SMP interconnection services and obligations for the world of the NGN. What Ofcom has instead done for now is to set in motion the consultation processes that can be used to try to provide clarity at such time as the issues are ripe for resolution.

For the most part, these processes represent well-conceived and appropriate attempts to encourage and enable industry to sort out the details of the transition, while reserving overall strategic direction to Ofcom. Many of these processes represent positive models of transparency and openness that Germany might wish to emulate; however, Germany would be well advised to implement a *single* umbrella consultation organization, with working groups to deal with specific issues, rather than multiple organizations with potentially overlapping responsibilities.

Wherever industry can reach consensus, the BNetzA should give great weight to industry views. Where industry cannot reach consensus, the BNetzA must be prepared to step in to resolve matters. BNetzA should be careful to avoid setting expectations for a consultative body that exceed its authority, or that exceed reasonable expectations of what a consultative body could realistically be expected to achieve.

## 4.2 The U.S.

The United States has not adopted the rhetoric of NGN, but for the most part the same technology and industry evolution is well advanced there.

The U.S. regulatory environment has been patched repeatedly to try to deal with the challenges of converged networks, to the point where it has become hopelessly unwieldy. At the same time, a number of U.S. ideas and practices on intercarrier compensation have achieved enviable results. The U.S. has achieved low or zero call termination rates, with a minimum of regulatory intervention. These low wholesale rates have resulted in correspondingly low rates and in flat rate plans at the retail level, which in turn have driven much higher mobile utilization than in Europe.

The high utilization and related consumer benefits flow from low *retail* rates, and only indirectly from low *wholesale* charges. Low termination rates in Germany, or a migration to Bill and Keep arrangements (zero termination rates) could serve as a vital enabler to low retail rates, in that low wholesale termination charges would make it possible for operators to offer lower-priced retail packages and/or flat rate plans. Low or zero termination rates are a potentially valuable means to an end; however, it should be possible to achieve these consumer benefits without duplicating the U.S. system in all its details.[78]

Of particular relevance are the FCC's still unrealized attempts to eliminate call termination mandates altogether. In effect, all telecommunications would follow the Internet-based example of interconnection based on commercial negotiations, with little or no regulatory intervention. The mobile market in the United States already operates quite successfully without mandated call termination fees, and achieves enormous utilization of mobile services. In the long run, it is quite possible that this is the only way to avoid economic distortions in a converged world. This approach merits serious consideration here in Germany.

The FCC's inability to implement a comprehensive migration to Bill and Keep is relevant to Germany primarily to the extent that it implies that existing operators might lobby intensively if they were to perceive themselves to be net losers in any change to interconnection arrangements. Whether a migration to low or zero termination charges would be realistically feasible in Germany has to be understood, then, in the context of the German regulatory and political system.

U.S. experience in trying to evolve the peering system to incorporate differentiated Quality of Service is also potentially of great relevance to Germany. The economic and business factors that inhibited progress in the U.S. would seem to be equally applicable here. There is no proven and demonstrated recipe for success. At the same time, it may

---

**78** How exactly this might most appropriately be achieved in Germany is beyond the scope of this study.

be possible to make cautious progress. For now, BNetzA should maintain its dialog with industry, and be alert to any opportunities that might emerge to move the process forward.