

Interconnection in the Internet Protocol (IP) Era

J. Scott Marcus

18-19 March 2013

Bingu Wa Mutharika International Centre
Lilongwe - Malawi

Interconnection and the changing electronic communications environment

- Traditional interconnection and IP interconnection
- Changes in the Use of Electronic Communications
 - Convergence
 - The rise of the mobile network
 - Fixed/mobile substitution
 - The rise of Next Generation Networks
 - NGN access
 - NGN core
- Forms of IP-based interconnection
 - Internet interconnection
 - Mobile IP Interconnection
 - IP-based interconnection of voice services

Interconnection in an evolving world

- Global attention over interconnection as networks evolve from traditional switched networks to IP-based networks.
- How this change is viewed, and the degree to which it is disruptive, can vary greatly
 - from one country to the next, and
 - from one regulatory system to the next.

Interconnection in an evolving world

- Switched fixed and mobile networks.
 - Regulation to address market power.
 - Termination fees in the absence of regulation will tend to be very high, for both large and small operators.
- Internet.
 - Peering: two providers exchange traffic only for their respective customers, often (but not always) with no explicit charges.
 - In most countries, no regulation of peering.
 - Peering arrangements are typically “Coasean”.
- What happens “when worlds collide”?

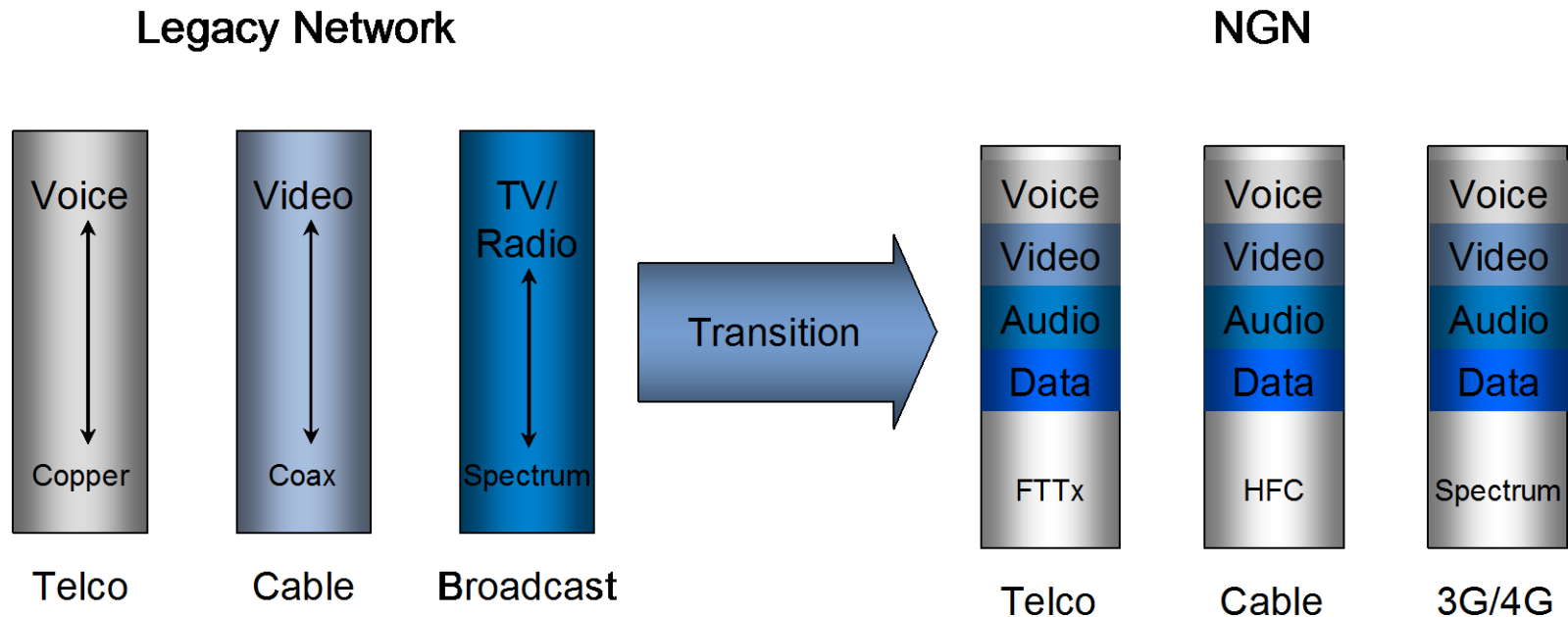
Traditional interconnection vs IP interconnection



Next Generation Networks (NGNs)

- Throughout the world, public networks are evolving into *Next Generation Networks (NGN)* based on the *Internet Protocol (IP)*.
- Different approaches to financing, business models, the role of government, and the regulatory approach are visible in different countries.

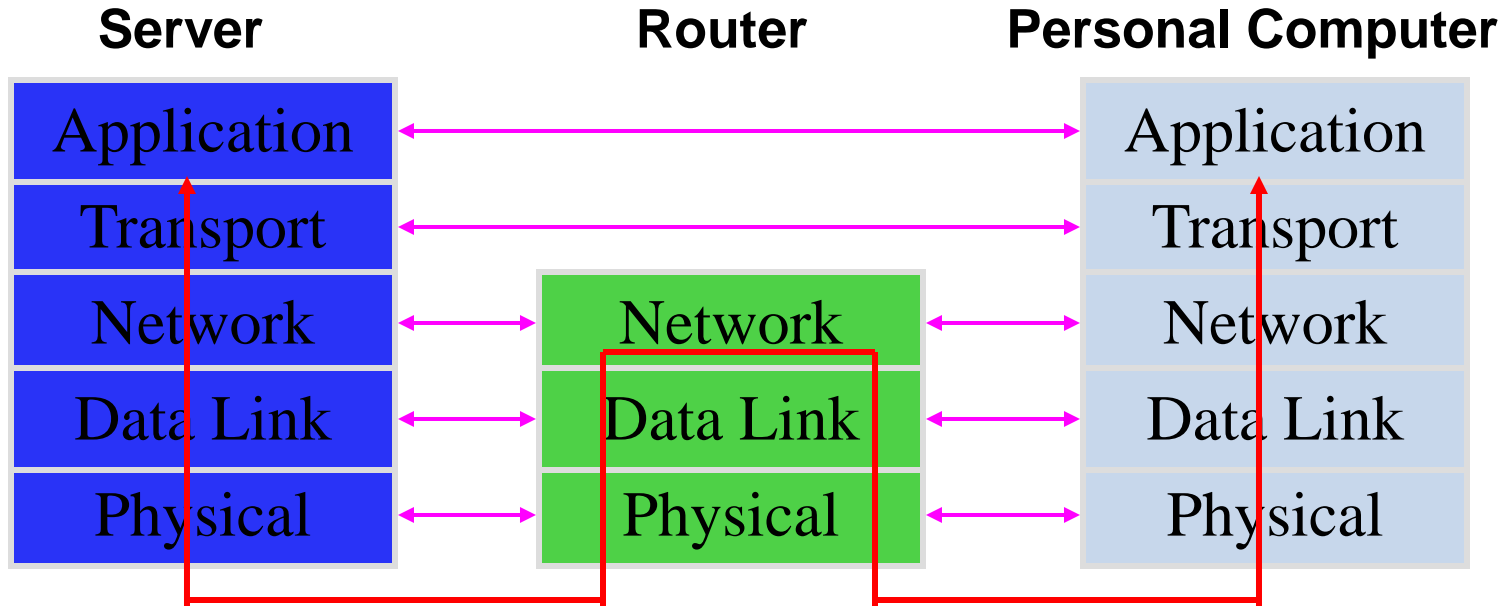
Convergence



- Historically, many networks delivered a single service.
- Today, any network can deliver (nearly) any service.

Basics: The TCP/IP Reference Model

Layers interact with peer layers

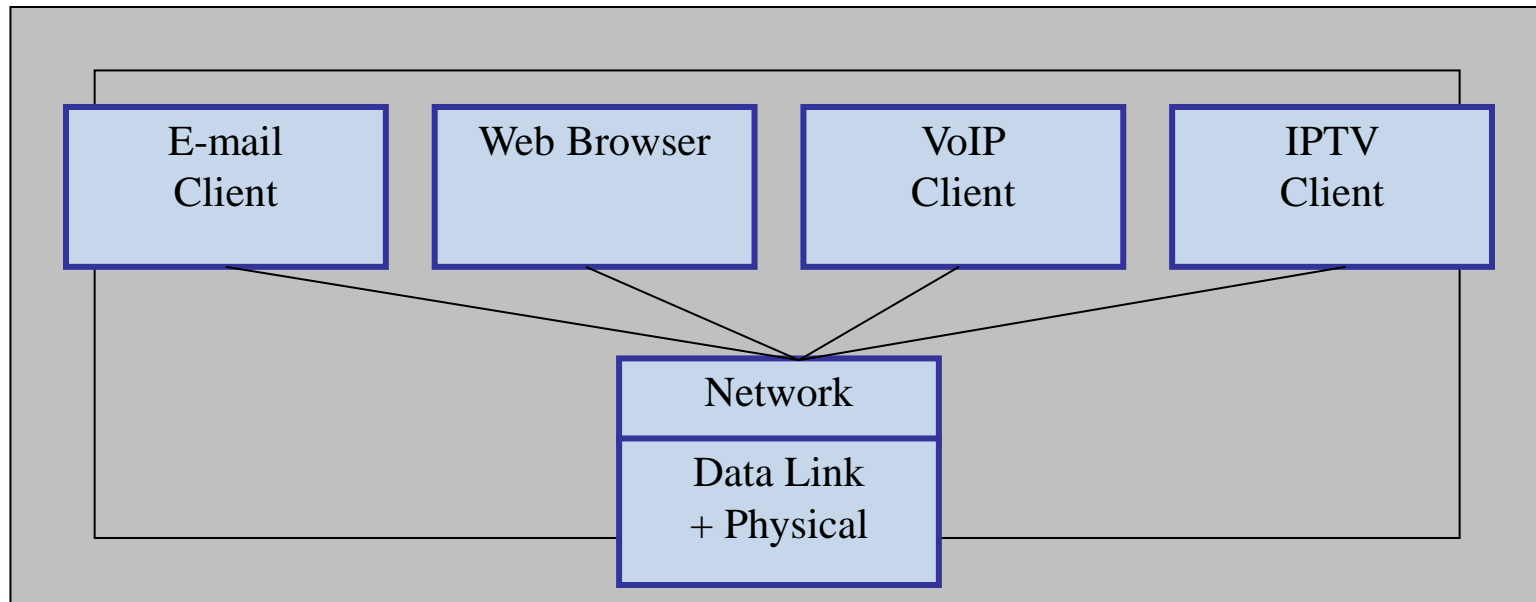


Layers derive services from successively lower layers

Basics: The TCP/IP Reference Model

- Physical Layer – the transmission facilities.
- Data Link Layer – the logical management of physical transmission facilities.
- Network Layer – forwarding and routing (Internet Protocol, or IP).
- Transport Layer – provides applications with datagram (UDP) or virtual circuit (TCP) services, as needed.
- Application Layer – provides services to the user.
 - Web
 - Email
 - VoIP
 - IPTV
 - Peer-to-peer file sharing

Basics: The TCP/IP Reference Model



Today, your broadband connection can support any combination of voice, video and data – provided that it is fast enough.

NGN Core, NGN Access

- Migration to IP has different drivers and different implications in different parts of the network.
- **NGN Core:** Replace traditional circuit switches with IP routers and VoIP gear in order to reduce operating cost and to offer new services.
- **NGN Access:** Drive fibre deeper into the network in order to offer higher speed, new services, greater reliability and lower operating cost.
- In developed countries, there has arguably been an excessive fascination with fibre-based NGA.

Achieving fibre-based NGN Access

Network Type	Cost per home accessed [in €]					
	DE	FR	SE	PT	ES	IT
VDSL	457	n.v.	352	218	254	433
PON	2,039	1,580	1,238	1,411	1,771	1,110
P2P	2,111	2,025	1,333	1,548	1,882	1,160

Source: WIK report on Next Generation Access for ECTA (2008)

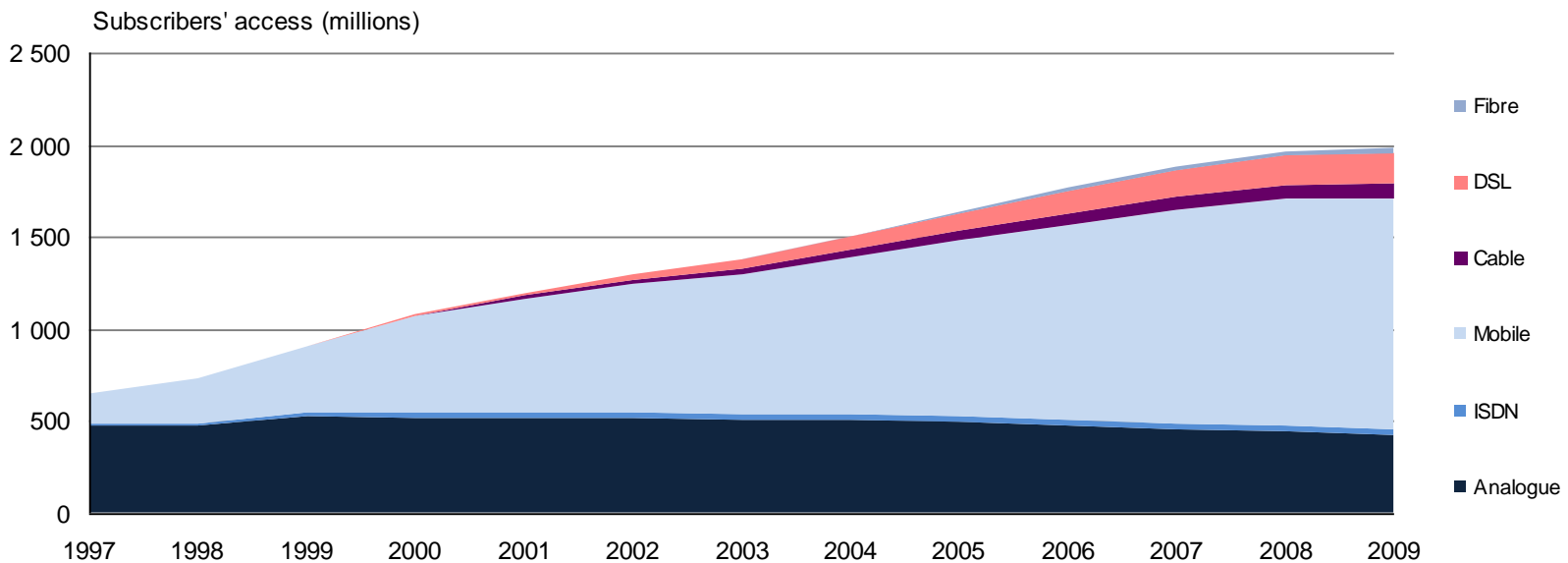
- Few countries are likely to achieve full NGA coverage without public stimulus/subsidy.
- Only limited prospects of replicating infrastructure.

Changes in the nature of competition

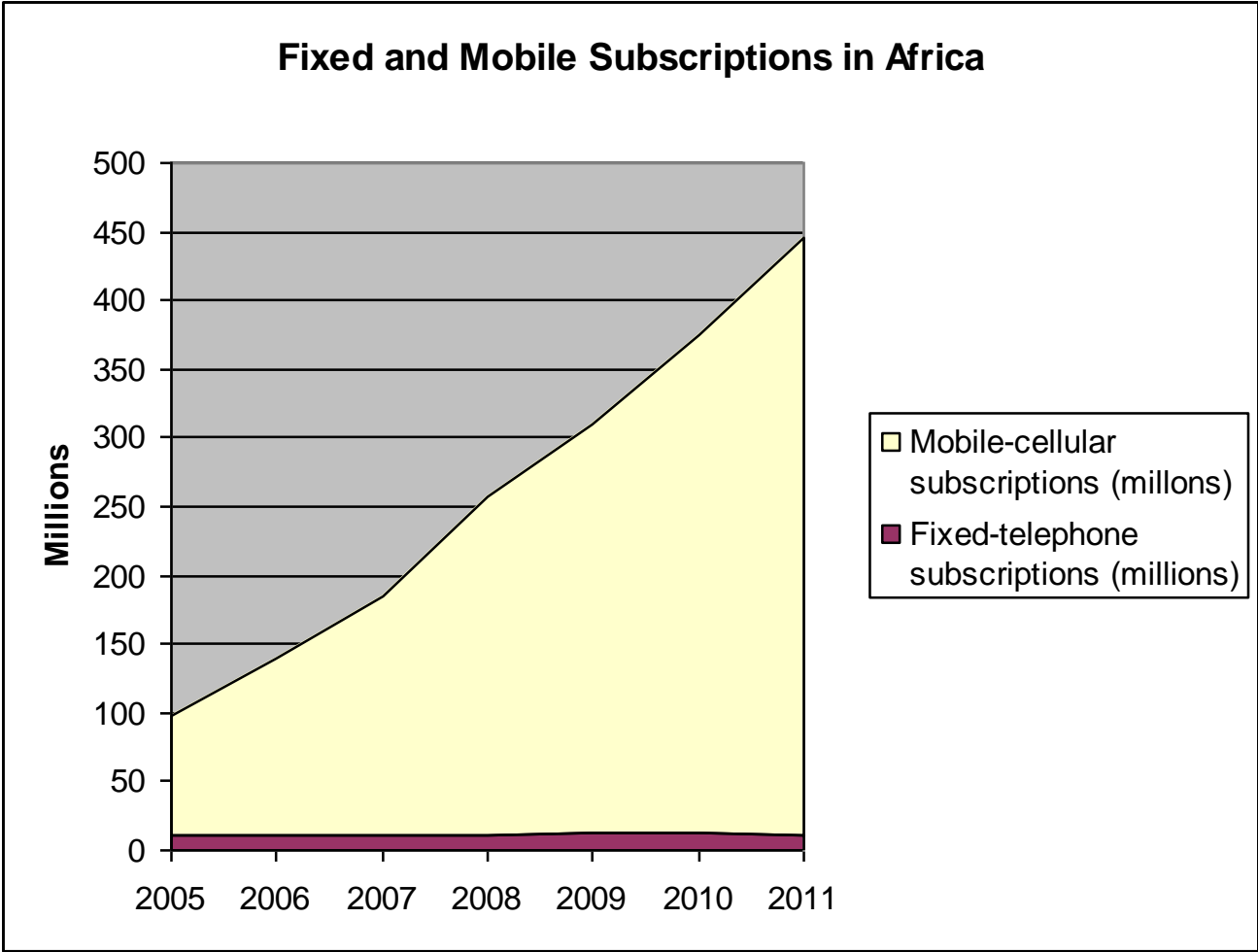
- Much of regulation and public policy of electronic communications deals with market power (SMP).
- The shift to NGN implies changes in the value chain, and thus subtly alters market power.
- Implications for regulation?
 - New forms of competition emerge?
 - Old barriers remain?
 - New barriers emerge?

We live in an increasingly mobile world

- In developed countries
 - Fixed lines are declining.
 - Mobile lines are increasing.
- In developing countries where the fixed network was never fully deployed, mobile growth is even more significant.

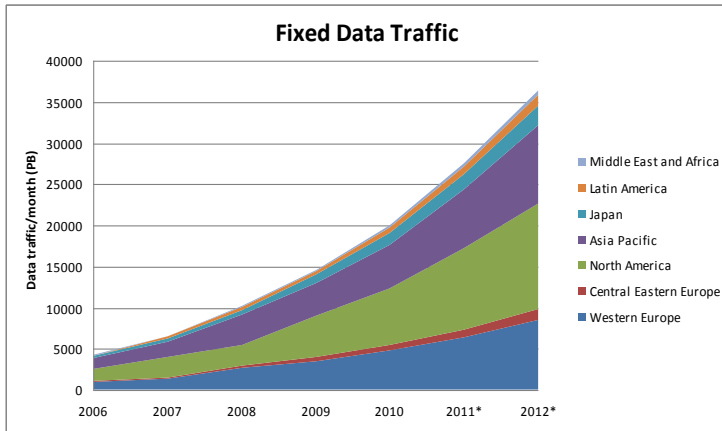


We live in an increasingly mobile world

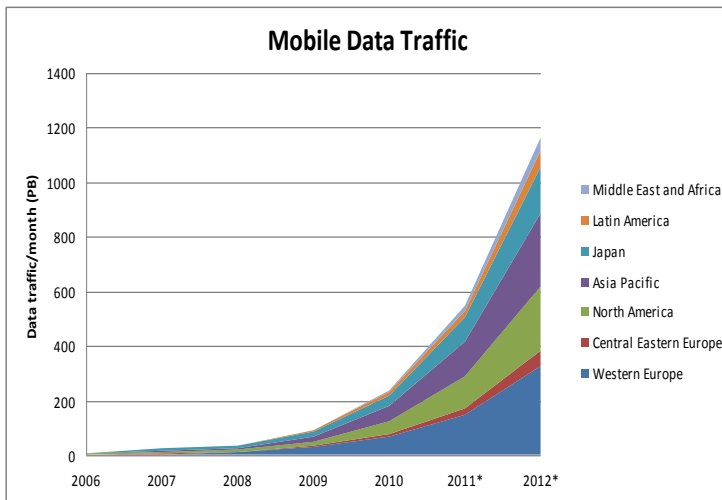


Data source: ITU, own calculations

Growth of fixed and mobile Internet traffic



Traffic is indeed increasing in both the fixed and the mobile networks.



Source: Cisco (2011), WIK calculations.

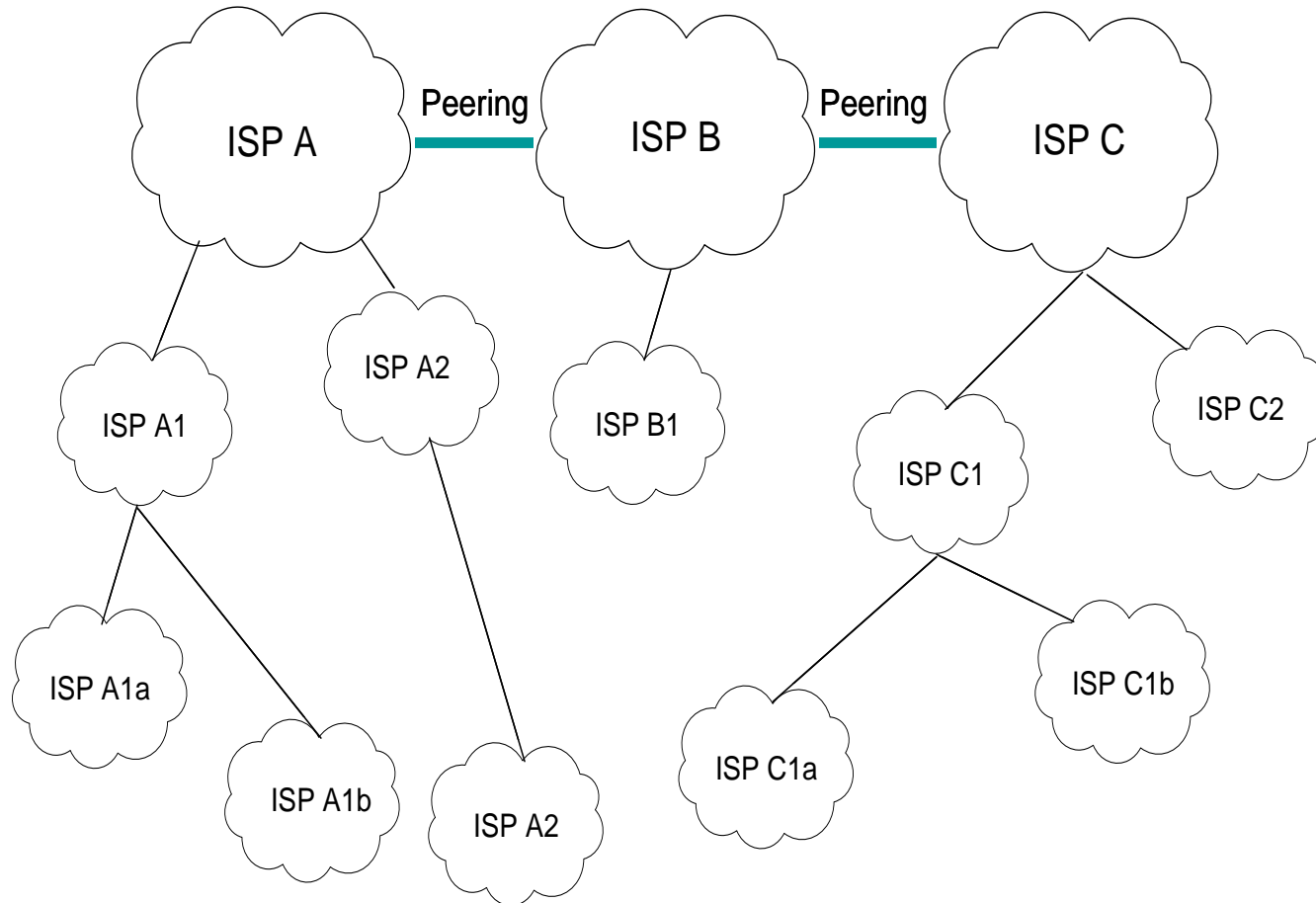
The emergence of IP-based interconnection

- Different forms address different needs.
- Internet interconnection
 - Widespread
 - Peering, transit, and variants on these models
- NGN IMS interconnection
 - Enables application level interconnection
 - No indication that this is happening
- Mobile IP Interconnection
 - GSMA GRX/IPX for data interconnection
- IP-based interconnection of voice services
 - Interesting developments

Peering, transit, and Internet access

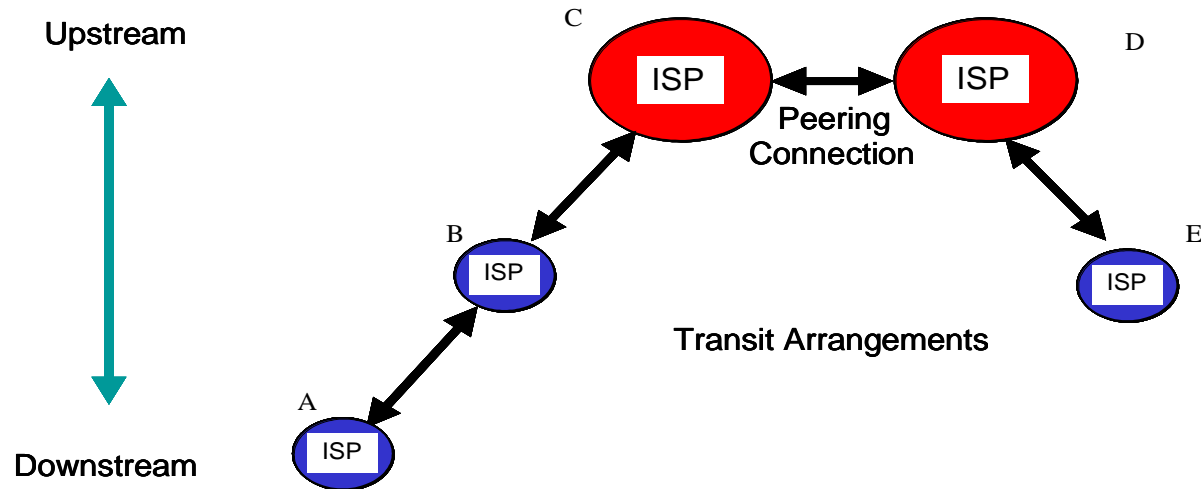
- Transit
 - The customer pays the transit provider to provide connectivity to substantially all of the Internet.
 - Essentially the same service is provided to consumers, enterprises, ISPs, content provider or application service providers.
- Peering
 - Two ISPs exchange traffic of their customers (and customers of their customers).
 - Often, but not always, done without charge.
- Variants of both exist.

Peering, transit, and Internet access



Peering, transit, and Internet access

Two peers and their respective transit customers



<i><u>Parties</u></i>	<i><u>Interconnection Arrangement</u></i>	<i><u>Typical Nature of Agreement</u></i>	<i><u>Typical Commercial Arrangements</u></i>
A – B	Transit	Bilateral	Payment reflects capacity, and may reflect volume of traffic or near-peak traffic level.
B – C	Transit	Bilateral	
E – D	Transit	Bilateral	
C – D	Peering	Bilateral	Often done without payment

Possible considerations for the CRASA region

- In developing countries, there can be big differences in the availability of high capacity international bandwidth.
 - Plentiful in capital cities near the coast
 - Cable landings and landing stations may pose a bottleneck.
 - Bandwidth may be scarce in the interior.
- Internet Exchange Points (IXPs) and informal arrangements.
 - Most peering arrangements are informal (Woodcock/Weller)
 - “Private” peering between major players may however represent a large fraction of interconnection bandwidth.
 - Some African countries have been leaders in implementing IXPs, while many others lag.
 - Market players may refuse to peer domestically.
 - Leads to inefficient “tromboning”.

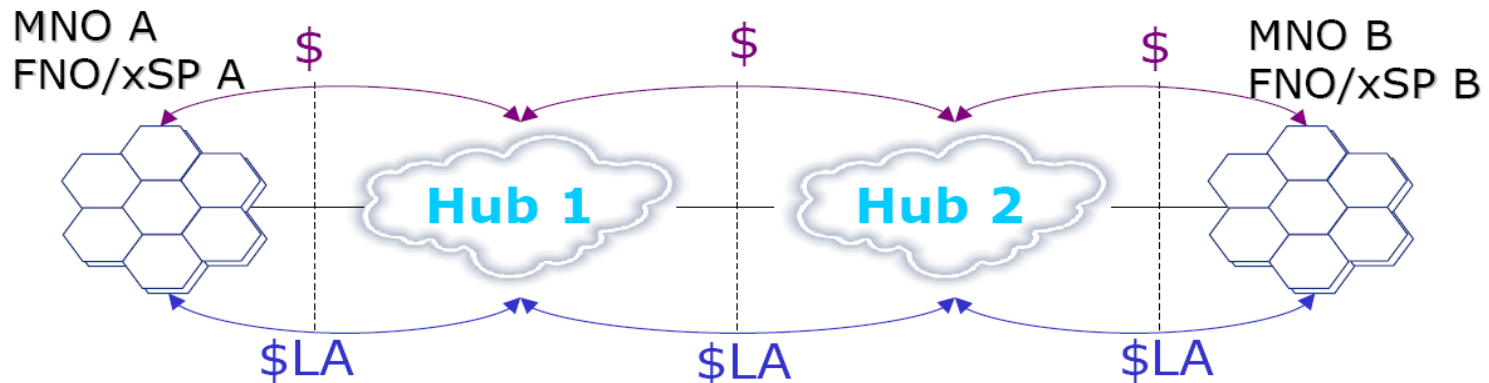
Mobile IP data interconnection

- Evolution of GSM-A interconnection architecture from GRX to IPX
 - Migration from basic peering model to QoS-oriented approach
 - IPX internals will be on “public” IP addresses, but will not be externally accessible.
 - Key voice infrastructure, including carrier ENUM and SIP servers, will thus be accessible only to MNOs or (possibly) to firms that accept MNO pricing and quality arrangements.
- Advantage: “Hubbed” architecture potentially overcomes transaction cost problems in getting QoS-capable services launched. (Although the value of differentiated QoS is not proven.)
- IPX with QoS capabilities could in principle be used for voice interconnection; however, we are not aware of it being used in this way.

Mobile IP data interconnection

Cascading QoS

- E2E SLA & Operational KPI's
 - Money Flows Along the Value Chain (Event Charging)
 - No free peering
 - Empowering all parties to secure SLA's



Copyright GSM Association

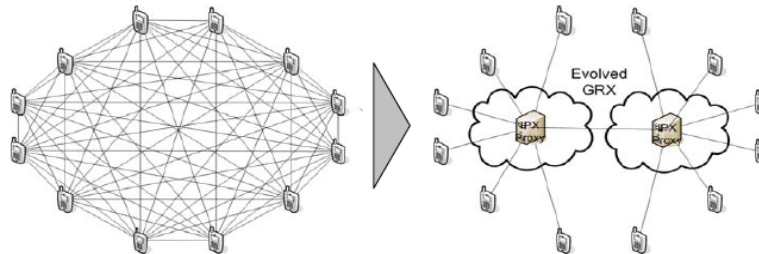
Mobile IP data interconnection

IPX Interconnect Models

IPX supports 3 different types of interconnect model

Interconnect Relationship	IPX Interworking Service	End to End QoS	Cascade Billing	Single Contract/ Connection
None - Internet		X		
Bilateral	Transport Only	✓	Termination billing takes place directly between Service Providers	Single contract with IPX provider but multiple contracts with connecting Service Providers
Bilateral	Service Transit	✓	✓	Single contract with IPX provider but multiple contracts with connecting Service Providers
Multilateral	Service Hub	✓	✓	✓

Multilateral interworking, or hubbing, is the principle benefit of the IPX. One contract many partners.



Copyright GSM Association

Voice Interconnection

- In most countries, call termination rates continue to be well in excess of true incremental cost, especially for the mobile network.
- Fixed incumbents, and even more so mobile operators, will likely resist external IP interconnection to their inherent voice service, except where charging arrangements mirror current arrangements.
- Tentative conclusion:
IP Interconnection \neq voice Interconnection

Voice interconnection

- Different levels of IP-based voice interconnection are visible in different countries.
- Little or no implementation, little discussion: most of the world.
- Active discussion, implementation only among competitors and/or cable operators: US, Canada, Spain, New Zealand, many more.
- Active incumbent implementation plans: Germany, Norway, Italy.
- Implemented by the incumbent: Denmark.

- This is not as far advanced as you might think!

Voice interconnection: Norway

- Telenor Norway plans to introduce IP interconnection to their voice services in 2013.
- Proposals for a technical solution (SIP-interconnect) and commercial terms have been presented to Telenor's wholesale customers.
- Concrete activities to develop a SIP interconnect offer:
 - Definition of standard connect products with IP VPN functionalities, managed routing, failover etc.
 - Implementation of a live testlab for initial testing and testing of new functionality.
 - Installation of dedicated SBCs for network connections.

Voice interconnection: Italy

- AGCOM put regulations in place.
- Ministry Economic Dev is developing technical standards.
- Telecom Italia published:
 - Reference Offer for IP interconnection on October 4, 2012
 - A manual of procedures for IP interconnection
- TI plan based on 16 VoIP Gateway Areas each of which is characterized by two Pols (geographical redundancy).
- Limited implementation to date (mostly Fastweb), but likely to increase in 2013 and beyond.

Voice interconnection: Germany

- Technologically conformant interconnection



- Technologically non-conformant interconnection

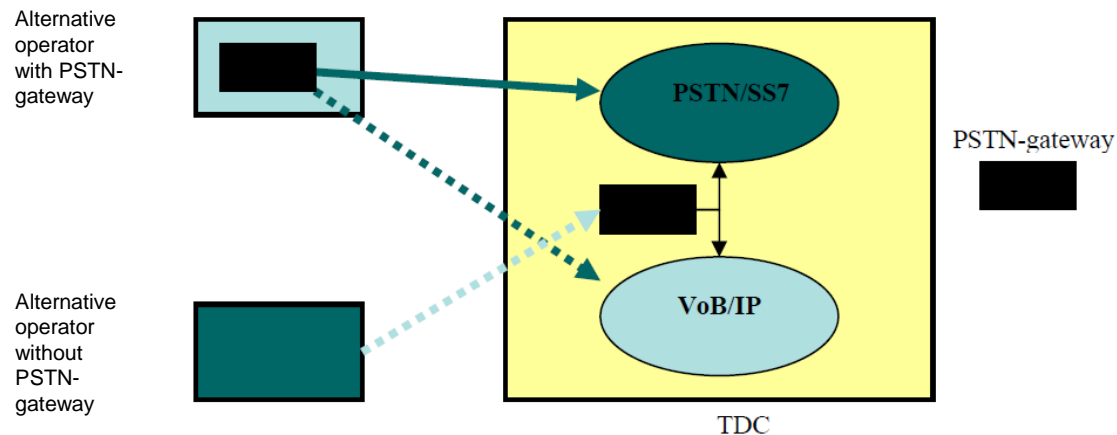


Voice interconnection: Germany

- No real deployment to date, but DTAG plans announced.
- Phase 1 Introduction (end 2011- mid 2013):
 - Introduction of NGN-IC
 - Test with every interconnection partner (ongoing!)
 - Parallel operation of PSTN-IC and NGN-IC
 - Market driven migration
 - PSTN continues to be in full operation
- Phase 2 Migration (mid 2013 – beginning 2015):
 - Tests with all interconnection partners must be completed before launch
 - Parallel operation of PSTN-IC and NGN-IC
 - Market-driven migration
 - Initiation of dismantling of PSTN
- Phase 3 Finalisation (beginning 2015 – end 2016).
 - Parallel operation of PSTN-IC and NGN-IC
 - PSTN products no longer marketed by Telekom Deutschland
 - PSTN Interconnection Points removed

Voice interconnection: Denmark

- TDC currently offers both TDM and IP interconnection to alternative operators.
- Alternative operators that have a PSTN gateway can connect either using IP or with traditional interconnection.



Voice interconnection: Denmark

- Following a review in 2007, NITA decided that cost models should evolve from circuit-switched to packet-switched.
- In 2008, both networks were implemented in the model.
- NITA has set prices based on the NGN model since 2010.
- Consistent with the principle of forward-looking costs, NITA does not reflect migration costs in its LRAIC modeling. Costs are modeled as if the efficient network were already rolled out.
- In 2011 an extensive update was carried out to the models. As part of this update a “pure LRIC” costing methodology was built into the model. However it was not used in the pricing decision for 2012, but is expected to be implemented in 2013.

- Migration of networks to IP is a global trend.
- IP interconnection for Internet traffic is well understood.
- As traditional voice services migrate to IP, there are good technical reasons to implement interconnection using IP.
- There is movement in this direction, and it is accelerating, but even in the most developed countries it is not as advanced as one might expect.

Economics of traditional and IP interconnection

- Key underlying economic principles
- Traditional network interconnection models
- Economic implications of migration to IP to interconnection

Economic background

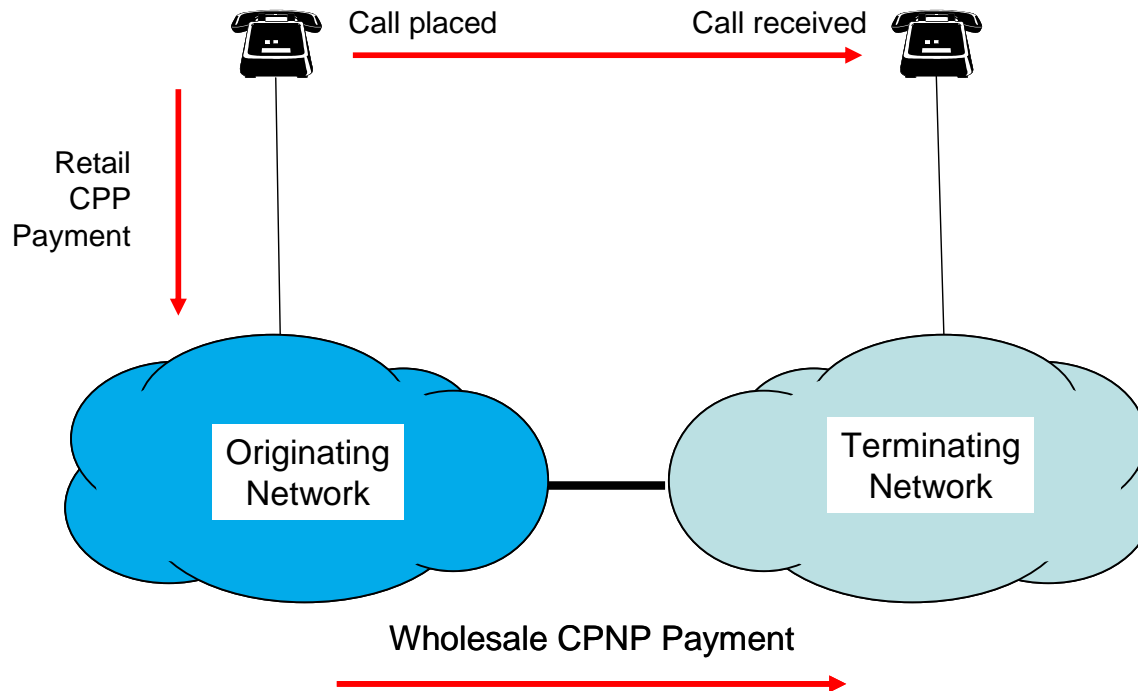
- Migration from switched fixed/mobile networks to NGN “marries” the traditional telephone network with the Internet. Which rules prevail?
- Switched fixed and mobile networks – regulated arrangements.
 - Regulation to address market power.
 - Termination fees in the absence of regulation will tend to be very high, for both large and small operators.
 - Lack of interconnection implies a connectivity breakdown.
- Internet – peering arrangements are typically “Coasean”.
 - Peering: two providers exchange traffic only for their respective customers, often (but not always) with no explicit charges.
 - Sharing of facilities costs for interconnection may be unequal.
 - In most countries, no regulation of peering.
 - Transit: usually interconnection for a fee to all destinations.
 - Lack of interconnection usually does not imply a loss of connectivity, but may have implications for costs.

Economic background: Retail arrangements

- Calling Party Pays (CPP)
 - Traditional arrangement based on presumed cost causality and presumed internalization of call externalities
- Receiving (Mobile) Party Pays (RPP/MPP)
 - Shared utilities from calls, receiver sovereignty.
 - True RPP systems are rare today.
- Flat rates: Calls included in monthly fees
 - Banded flat rates (buckets of minutes)
- Bulk of revenues comes from voice telephony; however, voice represents a sharply declining percentage of traffic.

Economic background: Traditional Fixed and Mobile Interconnection Models

- *Calling Party's Network Pays (CPNP)* wholesale arrangements



- An alternative (US, Canada, Hong Kong, Singapore) is to have negotiated arrangements under obligations of reciprocity, often resulting in no wholesale charges (Bill and Keep).

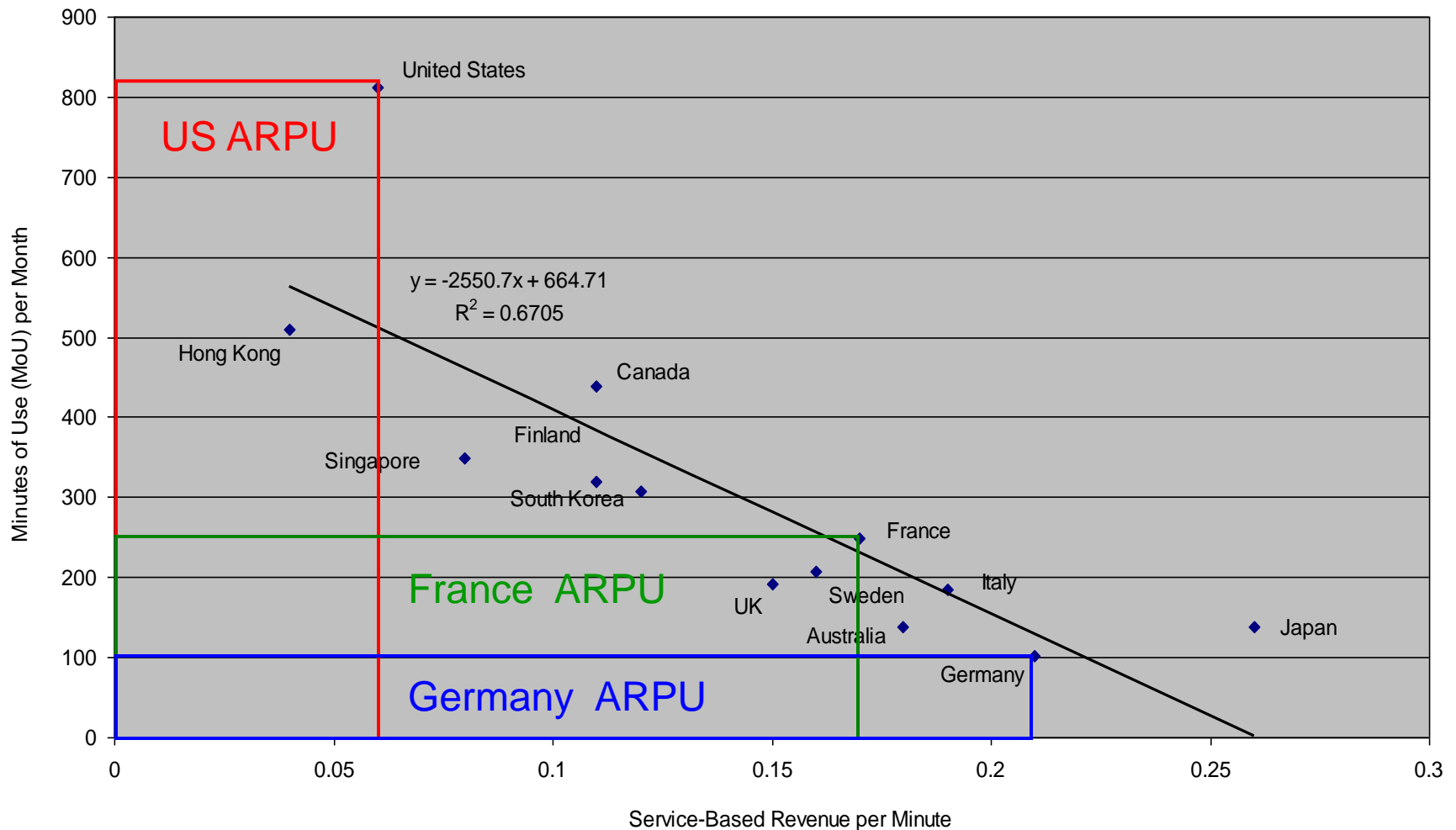


**Voice call
termination is a
tough problem!**

Economic background: How things work today: Wholesale and retail

- In an unregulated CPNP system, carriers will tend to establish very high termination charge levels (the *termination monopoly*).
 - Smaller operators will be motivated to set termination fees even higher than large operators.
 - The problem is addressed in the EU by regulating all rates.
- Several factors contribute to the termination monopoly.
 - Since the charges are ultimately borne by *another operator's customers*, normal market forces do not adequately constrain them.
 - Customers have no visibility into termination fees.
- Termination charges at the wholesale level interact with retail pricing arrangements.
 - The termination fee generally sets a floor on the retail price.
 - Where termination fees are high, they generally limit the applicability of flat rate or “buckets of minutes” plans.

SBR (retail price), Minutes of Use, and TRs



Economic background: How things work today: Wholesale and retail

- CPNP with high mobile termination rates tends to lead to:
 - Subsidies for mobile adoption, and thus rapid penetration.
 - High retail prices.
 - Exclusion of calls with high termination from flat rate plans.
 - Low usage.
- Rapid penetration can be beneficial, but not the rest.
- Wholesale and retail arrangements interact.
 - The TR sets an effective floor on retail usage-based price.
 - CPNP with high termination rates practically mandates CPP retail arrangements.
 - Low or zero termination rates, however, place no constraints on retail arrangements.

The regulatory solution

- Fixed network termination rates have been regulated in many countries for many years.
- Starting in 2003, the EU effectively regulated mobile termination rates (MTRs) as well.
- In 2009, the EU recommended that the rate for MTRs be determined by bottom-up modeling of “true LRIC” costs.
- There is a good argument that the MTR should reflect true costs.
- In the US, a complex system effectively sets the MTR to zero.
- MTRs averaged € 0.20 prior to regulation, but have declined steadily and are likely to fall to less than € 0.01 in most MS.

Economic background; Wholesale and retail

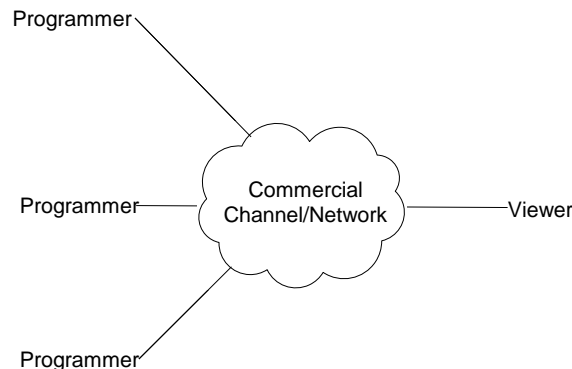
Region	Arrangements	Results
US	No fixed MTR, but obligations of symmetry and reciprocity	MTRs are zero, and FTRs are low.
Europe before 2003	No regulation at all for mobile	MTRs averaged €0.20
Europe today	Cost based MTRs on an accelerated glide path	MTRs dropping, heading toward perhaps € 0.005.

Two-Sided Markets

- A relatively new branch of economics deals with two-sided markets.
- In a two-sided market, a platform provider somehow benefits by bringing the sides of the market together.
- Payment could come from either side of the market; thus, relationships between price and cost that would be irrational in a conventional market might be reasonable in a two-sided market.
- Examples include broadcast television, and singles bars.
- Rochet, Jean-Charles/ Tirole, Jean (2004): Two Sided Markets : An Overview, March 200

Two-Sided Markets

- Broadcast television is a common and pertinent example.
- Payment comes from programmers/broadcasters, and ultimately from advertisers.
- The consumer typically pays little or nothing.
- In a conventional market, it would be strange for consumers to pay less than the cost of the service, but in a two-sided market it can be rational.



Two-Sided Markets

- Cable television provides a more complex demonstration of the dynamics of two-sided (or multi-sided) markets.
 - High value content providers such as premium sports can typically demand high payments from the cable operator, i.e. the provider of the two-sided platform
 - Providers of content that is valued less, or that is valued by fewer end-users, may not be able to command high payments, or for that matter may need to pay the cable operator to have their content transmitted.
 - The results of the negotiation are heavily dependent on the relative bargaining power of the parties.
 - Payment may flow in some cases from the end-user to the content provider, typically through the cable operator.

Two-Sided Markets

- From an economic perspective, the fact that a bargaining game is involved is not necessarily a problem, nor is the relevance of bargaining power.
- The negotiated outcomes can be economically rational and efficient.

Two-Sided Markets and interconnection

- Every electronic communications market is in some sense a two-sided market.
- HOWEVER, in not every case will a two-sided market analysis provide information that was not available in a conventional analysis.
- If the *structure* of prices matters, and not just the *level*, then a two-sided analysis may matter. (Tirole and Rochet)
- Areas where two-sided markets have been considered:
 - Voice call termination payments
 - Arrangements between network operators and content providers, especially as regards Quality of Service

Voice call termination as a two-sided market

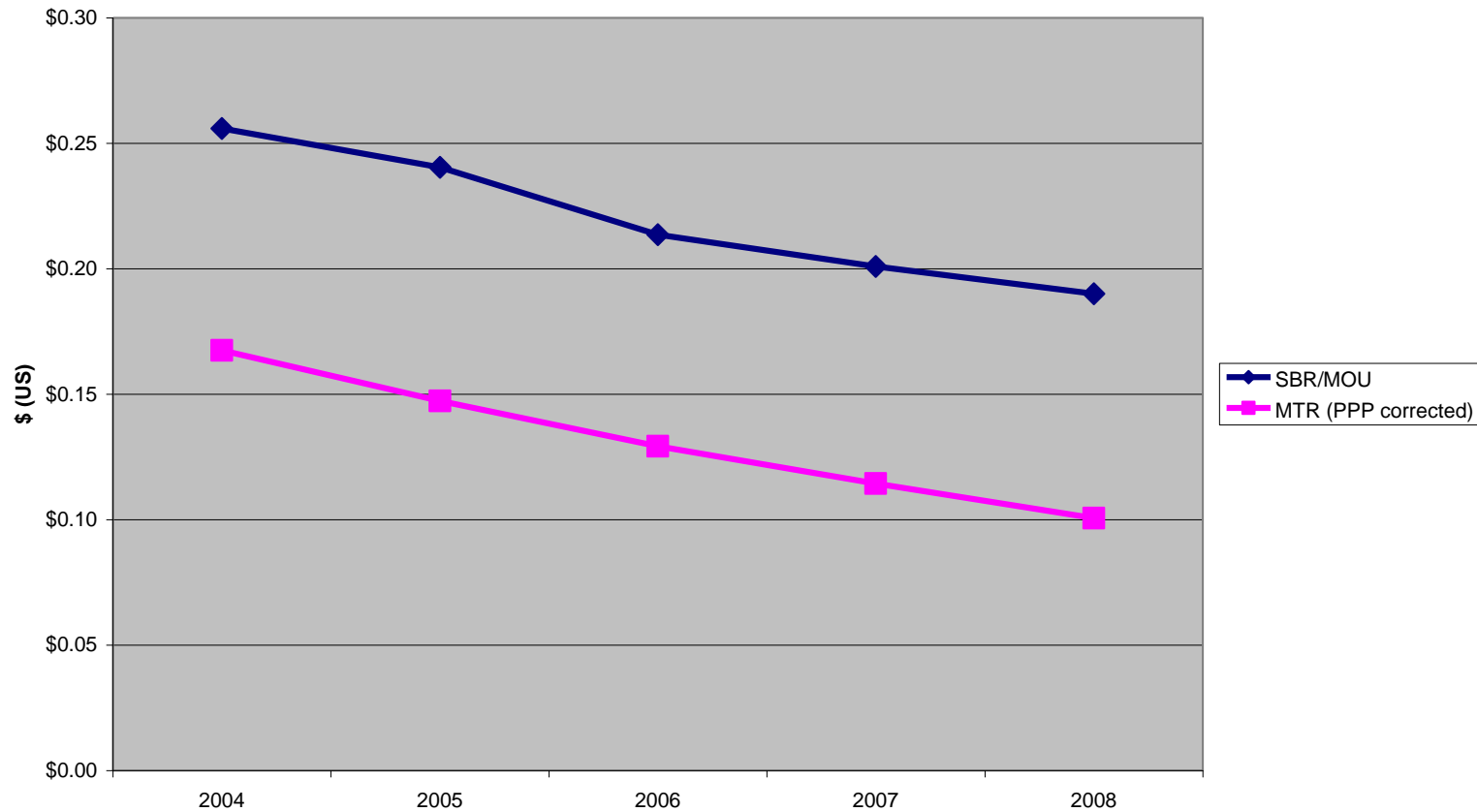
- The originating and terminating networks can collectively be viewed as a two-sided platform. Each can play either role.
- In this case, the called party is not very different from the party that placed the call, but the model is still valid.
- The terminating network possesses **terminating monopoly power** to the extent that no other network can complete calls to that number.
- *Calling Party Pays (CPP)* retail arrangements – where the receiving party nothing for the call – can be viewed as a two-sided market arrangement favouring call recipients over those who place calls.
- In reality, both parties benefit from the call. (Tirole and Jeon)

Voice call termination as a two-sided market

- *Calling Party Network Pays (CPNP)* wholesale arrangements – where the calling party's network make a payment to the receiving party's network – enable the two-sided market to work by funding the subsidy to the receiving party.
- What happens when the wholesale payment is reduced?
- It has been argued that firms will raise prices on the *originating* side to make up for lost *termination* revenues.
- Is this happening?

Voice call termination as a two-sided market

Service-Based Revenue per MoU vs MTRs in Europe

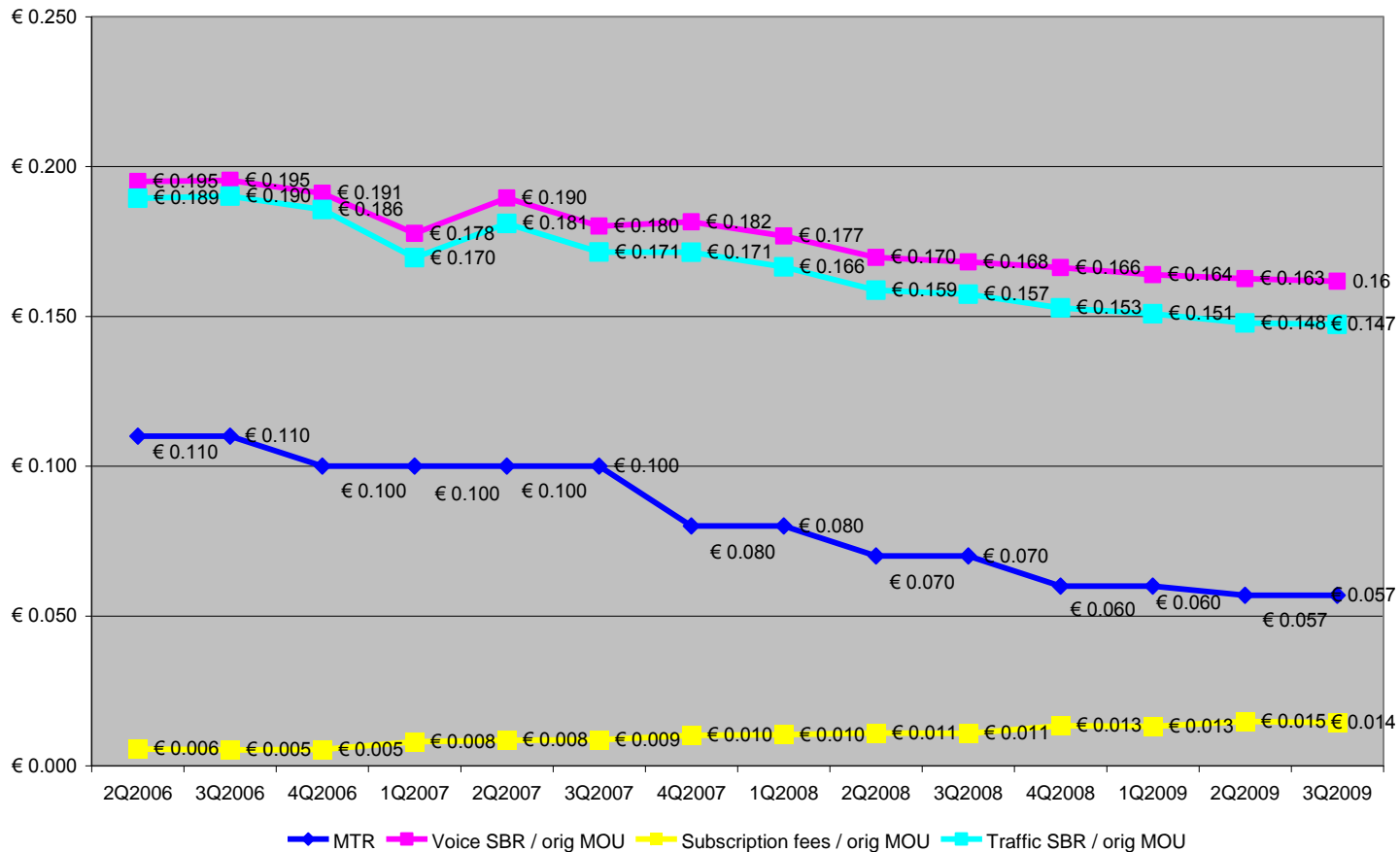


Voice call termination as a two-sided market

- A 2009 TERA study for the European Commission concluded that the relationship between MTRs and retail price is positive (the opposite from what one would expect based solely on two-sided market considerations).

	Panel data	Data Sources	Years	Results
Andersen and Hansen (2007)	9 European Countries	Ovum Wireless Intelligence	2003 - 2006	Cannot reject profit neutrality
Genakos and Valletti (2008)	OECD countries	Cullen International Teligen	2002 - 2006	Negative impact of MTRs on retail prices (stronger with competition and penetration)
Veronese and Pesendorfer (2009)	39 European and other OECD countries	OfTel Teligen Merrill Lynch European Commission	2002 - 2007	Small positive impact of MTRs on retail prices (insignificant with debiased data)
Growitsch, Marcus and Wernick (2010)	16 European countries	ERG Merrill Lynch CMT	2003-2008	Positive impact of MTRs on retail prices
Littlechild (2006)	44 countries	Merrill Lynch	2005	Insignificant impact of MTRs regime on retail prices

Voice call termination and retail revenues



Source: Spanish CMT (data), WIK calculations

Voice call termination as a two-sided market

- Two-sided market considerations may be relevant here, but many factors interact in complex ways.
- Wholesale revenues represent only 15% of total mobile revenues, so the direct impact is relatively small. This limits the magnitude of any two-sided market effects.
- For *Mobile-to-Mobile (M2M)* calls, the termination rate is a *cost*. Lowering the cost of calls in a competitive market tends to *reduce* the price, not to increase it.
- The fraction of calls that are M2M is increasing.
- Lowering MTRs limits the ability of large MNOs to prevent small ones from competing on price through on-net off-net price discrimination, thus enhancing competition.

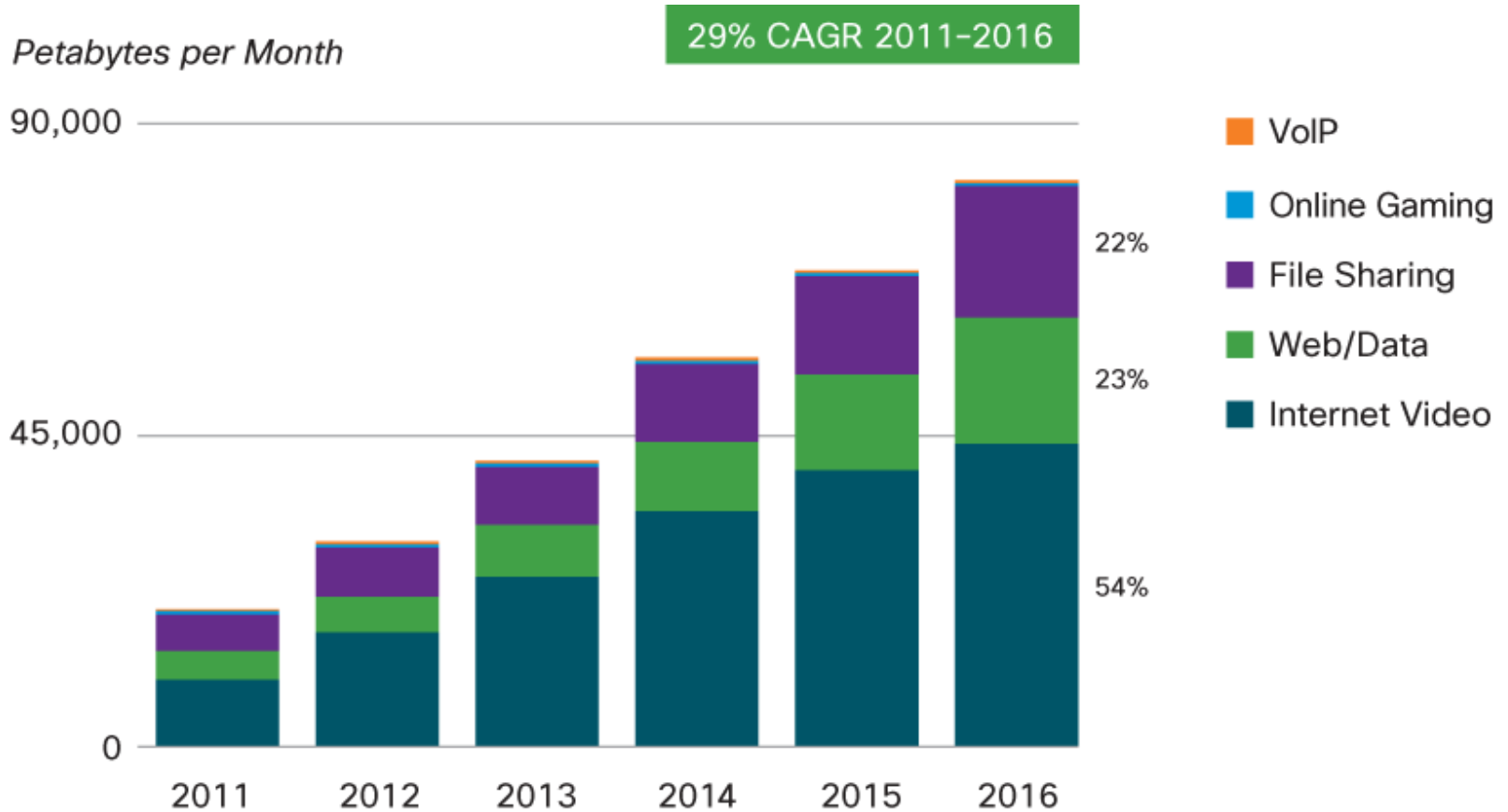
On-net Off-net price discrimination

- Most MNOs outside the US charge less at retail for calls to their own customers (on-net calls) than for calls to other MNOs.
- For on-net calls, the MNO faces the real marginal cost of termination, not the (possibly inflated) MTR.
- On-net off-net price discrimination favours MNOs with large shares of subscribers.
 - For a “large” MNO, many calls remain on-net.
 - For a “small” MNO, most calls must go off-net, and therefore face the high MTR as an incremental cost.
- For any MNO, it is challenging to price below the MTR. The more you sell, the more you lose.
- A low or zero MTR enables competitors to price aggressively, and constrains on-net off-net price discrimination.
- Largely absent in the US, and disappearing in Europe.

Challenges implied by migration to IP

- Network costs are driven by capacity requirements. In future integrated IP-based networks, where voice may represent only a small fraction of the traffic, total costs might have little to do with minutes of voice use.
- Traditional interconnection arrangements historically represented an attempt to use *wholesale* payments (between network operators) to correct for imbalanced *retail* payments (between service providers).
- To the extent that the network and service providers are different firms, and to the extent that voice is only a small fraction of the cost of the network, this system makes even less sense going forward than it did in the past.

The relative weight of VoIP and Internet video



Online gaming and VoIP forecast to be 0.73% of all consumer Internet traffic in 2016.
 Source: Cisco VNI Global Forecast, 2011-2016

- In conventional networks:
 - TRs would be very high absent regulation.
 - Regulated TRs should approximate long run incremental costs.
 - Those costs are declining rapidly due to IP-based technology.
- In Internet networks:
 - Transit prices are generally set by the market.
 - Peering charges, if any, are set by commercial negotiations.
 - Probably the best way to do it (but not everyone agrees).
- As voice moves to IP:
 - IP-based voice termination cannot have a substantially different cost from that of traditional voice.
 - Basing the TR on the IP-based cost seems to make sense.
 - Charges for PSTN-to-IP conversion could make sense.

Voice interconnection: Denmark

- Following a review in 2007, NITA decided that cost models should evolve from circuit-switched to packet-switched.
- In 2008, both networks were implemented in the model.
- NITA has set prices based on the NGN model since 2010.
- Consistent with the principle of forward-looking costs, NITA does not reflect migration costs in its LRAIC modeling. Costs are modeled as if the efficient network were already rolled out.
- In 2011 an extensive update was carried out to the models. As part of this update a “pure LRIC” costing methodology was built into the model. However it was not used in the pricing decision for 2012, but is expected to be implemented in 2013.

Quality of Service (QoS), Net Neutrality, implications

- Differentiated Quality of Service (QoS) for IP-based traffic
 - Technical aspects
 - Applications requirements
- Network Neutrality debates
 - Definitions of net neutrality
 - Technical aspects
 - Economic aspects of price/quality differentiation
 - Comparisons between Europe and the US
- QoS and the debate between content providers and network operators
- Concluding remarks

What is QoS in an IP-based NGN?

- In traditional fixed telephone networks, there tended to be a great deal of concern about blocking probability, much less about voice quality once the call was allowed to complete.
- In an IP-based system under load, packets can routinely be queued for transmission, or dropped if the queue is too long.
- These delays are not a failure mode – they are a normal aspect of Internet Protocol operation.
- What factors are most critical to the Quality of User Experience in such an environment?

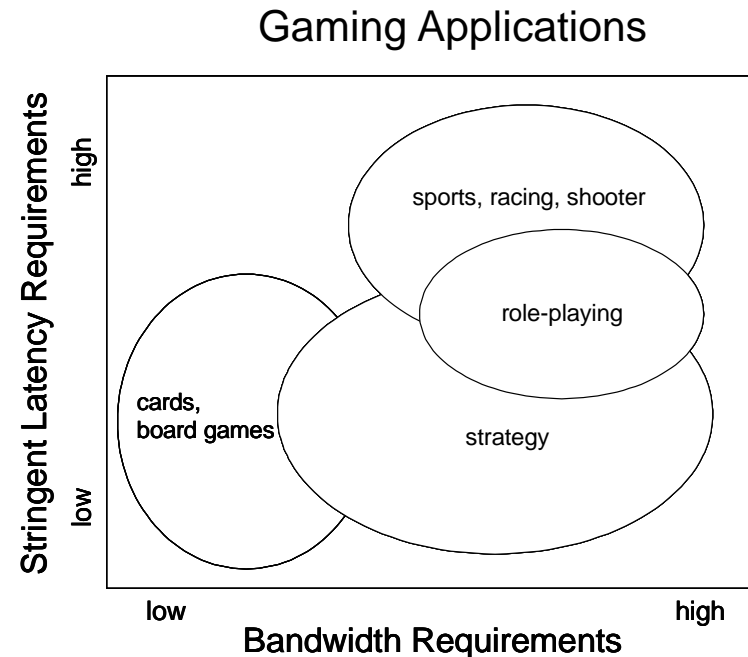
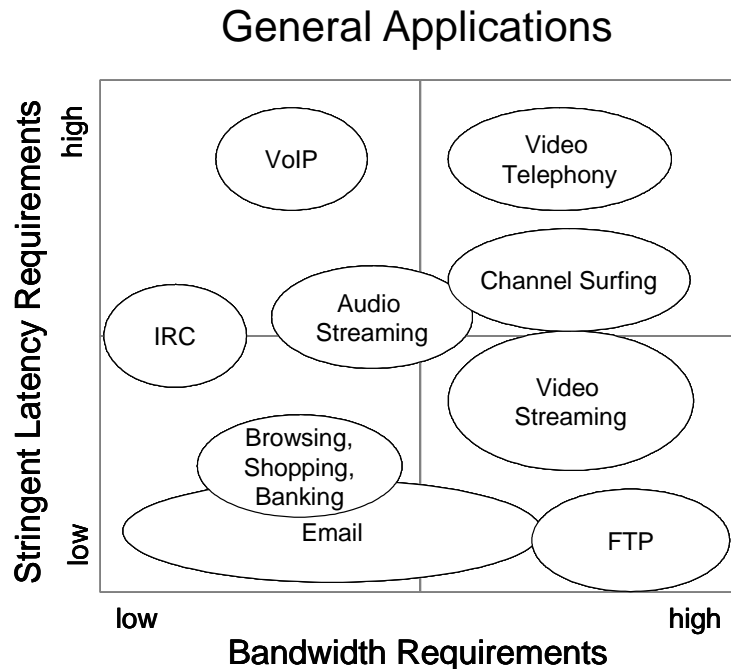
Quality of Service (QoS) and Quality of Experience (QoE)

- QoS parameters and mechanisms are important to enable network operators to design, build and manage their networks, but they are not directly visible to end-users.
- Crucial for end-users, however, is the quality that they personally experience during their use of a service.
- These Quality of Experience (QoE) requirements are strongly dependent on the application.
 - E-Mail has little sensitivity to packet loss and delay.
 - Real-time two-way Voice over Internet Protocol (VoIP) tends to be highly sensitive – delays more than some 150 msec cause problems.
 - One-way video may or may not be sensitive, depending on user expectations for how quickly the stream starts (zapping time).

What is QoS in an IP-based NGN?

- **Bandwidth:** the maximum number of bits that a transmission path can carry.
- **Propagation delay:** The time that a packet requires, as a function of the length of all transmission path and the speed of light through that particular transmission path.
- **Queuing delay:** The time that a packet waits before being transmitted. Both the average delay and variability of delay (jitter) matter, since the two together establish a confidence interval for the time within which a packet can be expected to arrive at its destination.
- **Packet loss:** The probability that a packet never reaches its destination. This could be due to transmission errors, but errors are quite rare in modern fibre-based fixed networks. More often, packets are lost because the number of packets waiting for transmission is greater than the available storage capacity (buffers).

Differentiated Quality of Service (QoS): Application Needs

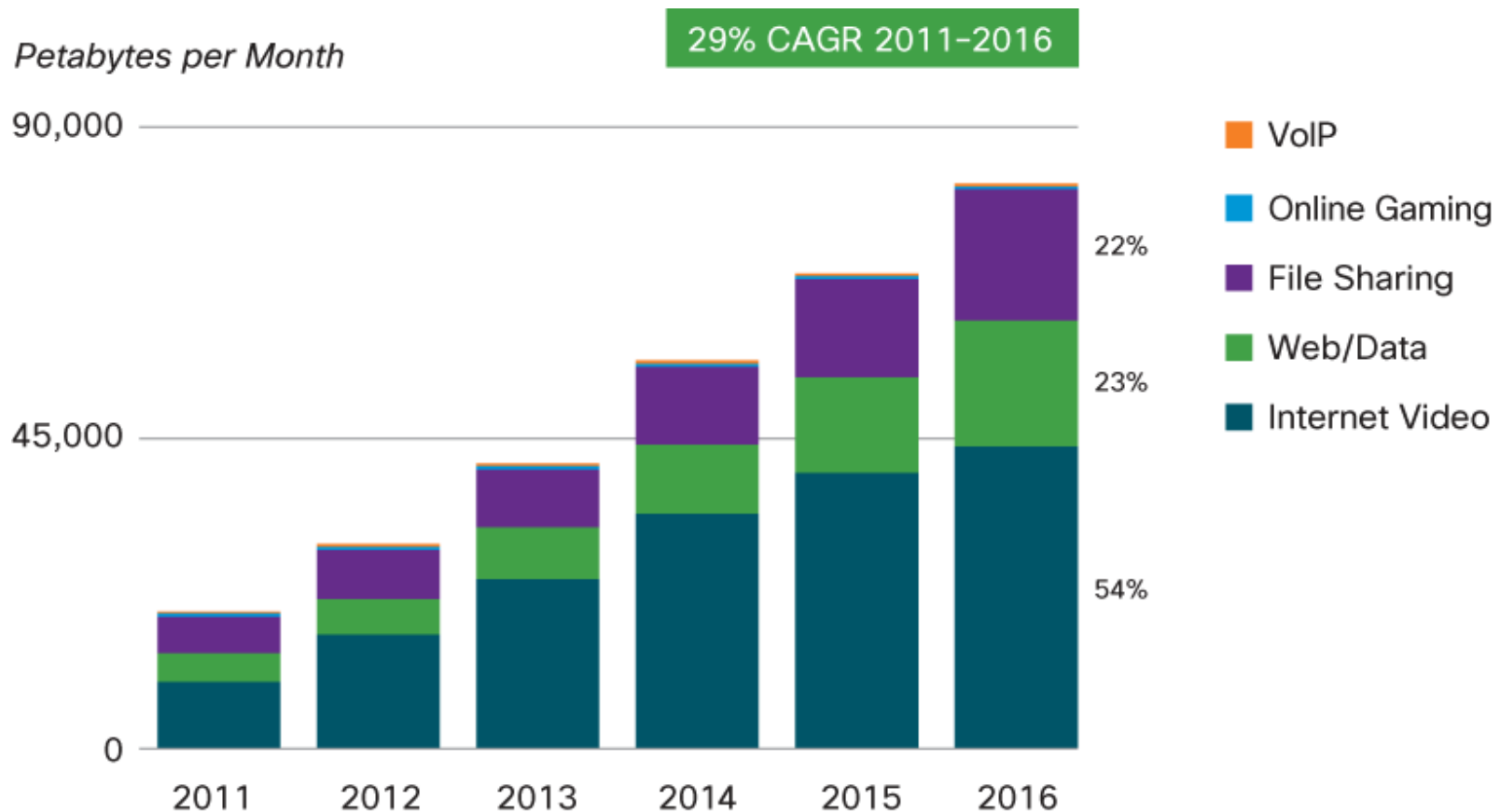


- Real time bidirectional audio: stringent requirements
- Email: liberal requirements
- Streamed audio and video: fairly liberal requirements (except where the impact channel surfing/"zapping time").

Differentiated Quality of Service (QoS): Application Needs

- For voice, if delay exceeds about 150 milliseconds, both sides may begin to speak at once.
- Not all video is delay-sensitive.
 - For real-time videoconferencing, similar considerations apply to delay; however, bandwidth requirements are far greater.
 - For streamed video, if it is permissible to wait a second or two at the outset, a jitter buffer can accommodate typical delays.
 - Interconnection is not relevant to all video. Much video is originated close to the end-user (within the end-user's own network, in order to save transmission costs).
- Certain interactive games may be highly delay-sensitive.
- Data applications tend to be less sensitive, but some (e.g. web-browsing) are more delay-sensitive than others (e.g. e-mail).

The relative weight of VoIP and Internet video



Online gaming and VoIP forecast to be 0.73% of all consumer Internet traffic in 2016.
 Source: Cisco VNI Global Forecast, 2011-2016

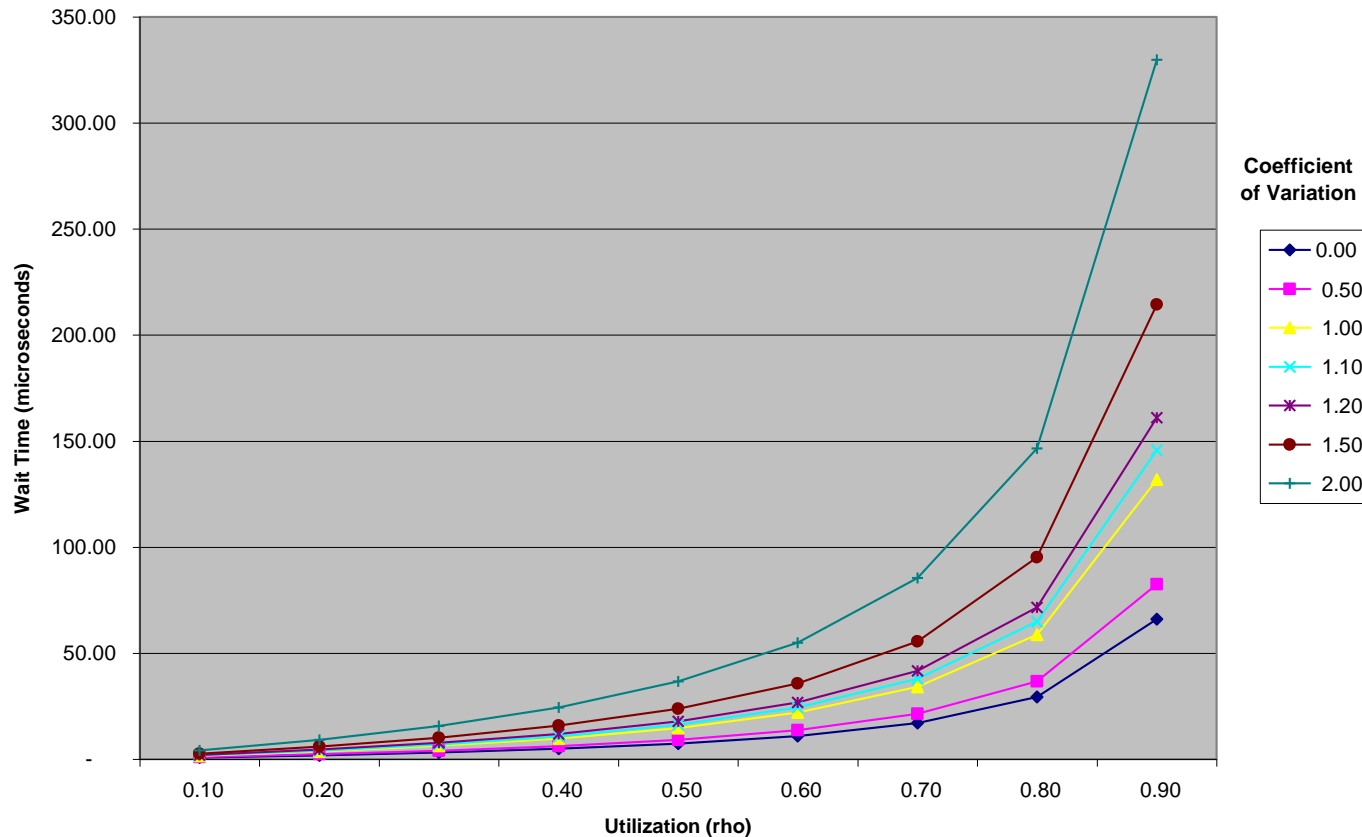
Differentiated Quality of Service (QoS)

- At a technical level, QoS is not fundamentally hard.
 - DiffServ is technically trivial.
 - MPLS in a single network is technically trivial.
 - Cross-provider MPLS is only marginally harder.
 - Even RSVP is not that hard. My former company, BBN, had working production RSVP-compliant networks in 1995!
- In terms of the basic economics, QoS is not fundamentally hard.
- Differentiated QoS *within* a network is, in fact, commonplace.
- Nonetheless, there is no significant roll-out of differentiated QoS *between* networks.

➤ **WHY NOT?**

Economic theory and QoS issues: Differentiated Quality of Service (QoS)

M/G/1 Queuing Delay (155 Mbps Link)



M/G/1 queueing analysis of the performance *of a single link*

(with clocking delay of 50 μ secs (284 byte packets) and a 155 Mbps link)

Differentiated Quality of Service (QoS)

- As we have seen, per-hop delay, even in a network with 90% load, is about 1,000 times less than the 150 millisecond delay “budget” for real-time bidirectional voice.
- IMPLICATION: Most of the time, and under normal conditions, variable delay in the core of the network is barely perceptible to the users of VoIP or other delay-sensitive applications.
- FURTHER IMPLICATION: Consumers will not willingly pay a large premium for a performance difference that they cannot perceive.
- Packet delay is more likely to be an issue:
 - For slower circuits at the edge of the network
 - For shared circuits (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

Implementing inter-provider QoS

- Although the technology is reasonably straightforward, little practical experience in enforcing QoS across IP-based networks.
- It is not due to a lack of standards – there are too many standards, not too few.
- Classic problem of introducing change into a technological environment:
 - Network effects – no value until enough of the market has switched (Cf. Jeff Rohlffs)
 - Long, complex value chains.
 - Costs and complexity of transition.
- Analogous problems have slowed IPv6 and DNSSEC.

Implementing inter-provider QoS

- Efforts to extend Quality of Service (QoS) across network operators have failed to catch fire for many reasons:
 - **Scale:** Bilateral peering arrangements will tend to be acceptable to both network operators only when the networks are of similar scale, or more precisely when both networks can be expected to be subject to similar cost drivers for carrying their respective traffic.
 - **Traffic balance:** Where traffic is significantly asymmetric, cost drivers are likely to also be asymmetric.
 - **Monitoring and management:** There are many practical challenges in determining whether each network operator has in fact delivered the QoS that it committed to deliver.
 - **Financial arrangements:** There has been no agreement as to how financial arrangements should work. In particular, there has been enormous reluctance on the part of network operators to accept financial penalties for failing to meet quality standards.

Implementing inter-provider QoS

- Many efforts over the years to define inter-provider QoS standards.
- One of the best and most practical was organised by MIT, with substantial industry participation.
- The following values from the MIT white paper would appear to be reasonable for IP interconnection suitable for real time bidirectional voice:

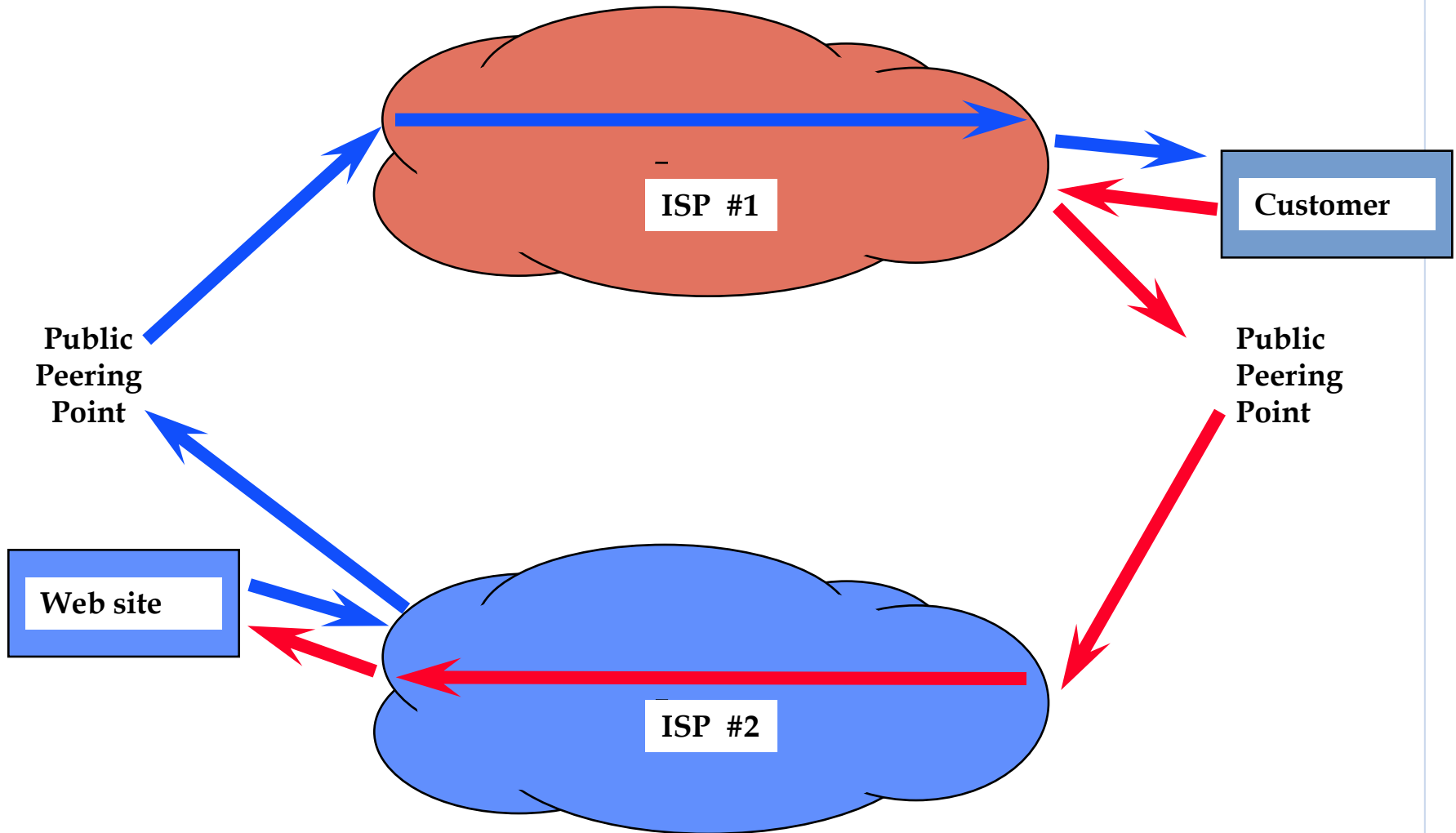
Delay:	100 msec
Delay Variance:	50 msec
IPPM Loss Ratio:	1×10^{-3} (One Way Packet Loss)

- The MIT WG white paper also explains how to measure these, and how to allocate end-to-end requirements to multiple networks. IPPM probes could be suitable.
- A challenge: No network operator will want another to operate probes within its network.

Implementing inter-provider QoS

- As part of the functional/operational separation of Telecom New Zealand, there were commitments
 - To interconnect with competitors using IP
 - To support a suitable QoS for VoIP in those interconnections
- The first of these is in place.
- For the second, Telecom New Zealand made a quite interesting proposal, based on their methodology for the first of these.

Peering and Shortest Exit



Implementing inter-provider QoS: TNZ offer

- Division of New Zealand into 29 interconnection areas;
- Willingness to interconnect with any network operator of any size (without settlement payments for IP traffic) to interchange data with TNZ customers within that interconnection area, provided that the access-seeking network operator has made arrangements to get its traffic to the interconnection area;
- Availability of IP traffic transit arrangements from TNZ at reasonable wholesale prices to get the traffic to the desired interconnection area;
- A fair process for achieving physical interconnection within an IP interconnection area if desired;
- Two classes of services offering performance better than “best efforts”; but
- No specific penalties or payments if traffic is delivered with quality less than that committed.

Implementing inter-provider QoS

- Telecom New Zealand proposal should be workable.
- Technically, it is nearly identical to the means by which some of the largest backbones interconnect globally. It differs only in geographic scale.
- Economically, it is similar to (apparently sound) proposals by Ingo Vogelsang and Patrick de Graba.
- Deals simply and elegantly with size differences among network operators.

Observations: Price and quality differentiation

- Not every application needs differentiated quality of service, but for those that do, QoS control makes technical sense.
- It has existed within IP-based networks for many years.
- It has failed to develop (much) between networks for reasons that have more to do with economics than with technology.

Network neutrality

- How should we define network neutrality?
- Why does net neutrality raise concerns?
- What harms are known to date?

How should we define network neutrality?

- Network neutrality has taken on various meanings:
 - The ability of all Internet users to access the content or applications of their choice.
 - Assurance that all traffic on the Internet is treated equally, whatever its source, content or destination.
 - Absence of unreasonable discrimination on the part of network operators in transmitting Internet traffic.
- These are not equivalent to one another!
- The use of various forms of quality differentiation for Internet traffic within networks has been routine for decades.
- This differentiation serves in most cases (but not necessarily in all) to benefit consumers.

Why does net neutrality raise concerns?

- Breaches of network neutrality have raised a range of different fears related to negative impacts on:
 - Competitiveness
 - Innovation
 - Freedom of expression
 - Consumer welfare generally
 - Privacy
- The issues that have been raised are subtle and complex.
- A simplistic, one-size-fits-all solution could potentially raise equally serious concerns.

What harms are known to date? /1

- The European Commission conducted a public consultation on network neutrality in late 2010.
- The Commission found a consensus among “...network operators, internet service providers (ISPs) and infrastructure manufacturers that there are currently no problems with the openness of the internet and net neutrality in the EU ... They maintain that there is no evidence that operators are engaging in unfair discrimination in a way that harms consumers or competition. This general view is supported by a number of Member States.”

What harms are known to date? /2

- There have been scattered complaints, some of them credible, of
 - mobile network operators (MNOs) blocking or charging excessive prices for VoIP
 - blocking or throttling of traffic such as file sharing
- BEREC considered the complaints relevant, but
 - they “... may not necessarily represent breaches of network neutrality”;
 - many were finally resolved “without any formal proceedings”;
 - and
 - the incidents “have not led to a significant number of investigations by NRAs”.

What harms are known to date? /3

- There appear on balance to be few if any documented, clearly problematic incidents in Europe to date.
- There is no demonstrated, sustained pattern of systematic and abusive discrimination.
- Despite all of this, possible concerns for the future remain.

Economic background of network neutrality

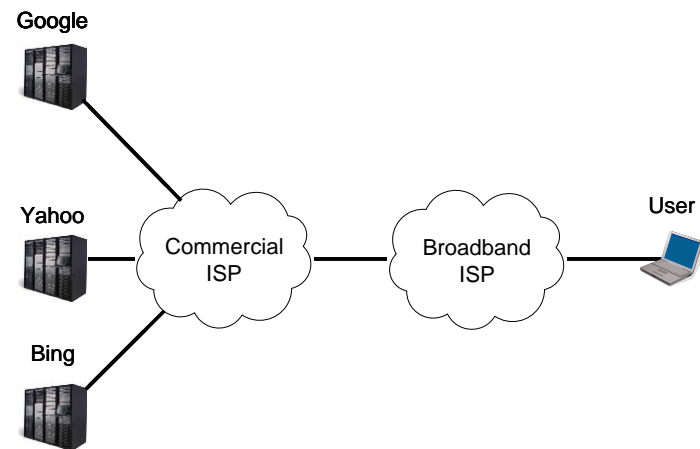
- Quality differentiation
- Economic foreclosure
- Two-sided (or multi-sided) markets

Price and quality differentiation

- Quality differentiation and price differentiation are well understood practices.
- In the absence of anticompetitive discrimination, price and quality differentiation generally benefits both producers and consumers.
- We typically do not consider it problematic if an airline or rail service offers us a choice between first class and second class seats.

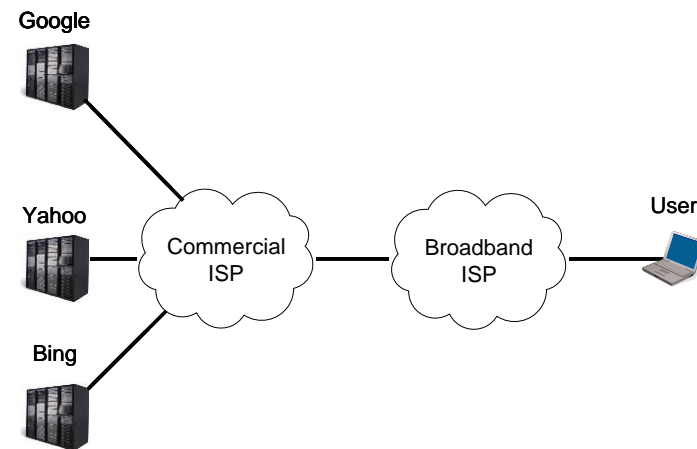
Economic foreclosure

- When a producer with market power in one market segment attempts to project that market power into upstream or downstream segments that would otherwise be competitive, that constitutes economic foreclosure.
- Foreclosure harms consumers, and imposes an overall socio-economic deadweight loss on society.



Two-sided markets

- The Internet can be thought of as a two-sided market, with network operators collectively serving as a platform connecting providers of content (e.g. web sites) with consumers.
- The ISPs *collectively* are the two-sided platform. Their interests are not aligned.
- Each side of the market pays its own ISP (or makes equivalent investments).



- Harmful deviations from network neutrality (especially economic foreclosure) tend to critically depend on the presence of market power.
- Harmful deviations are unlikely if:
 - Market power either is absent, or has been addressed through proper regulatory means, such that consumers have choice.
 - Consumers are informed.
 - Switching costs are not exorbitant.
- Regulations to enforce net neutrality would be hard.
- Better to make violations unlikely.

Differences between the US and the EU

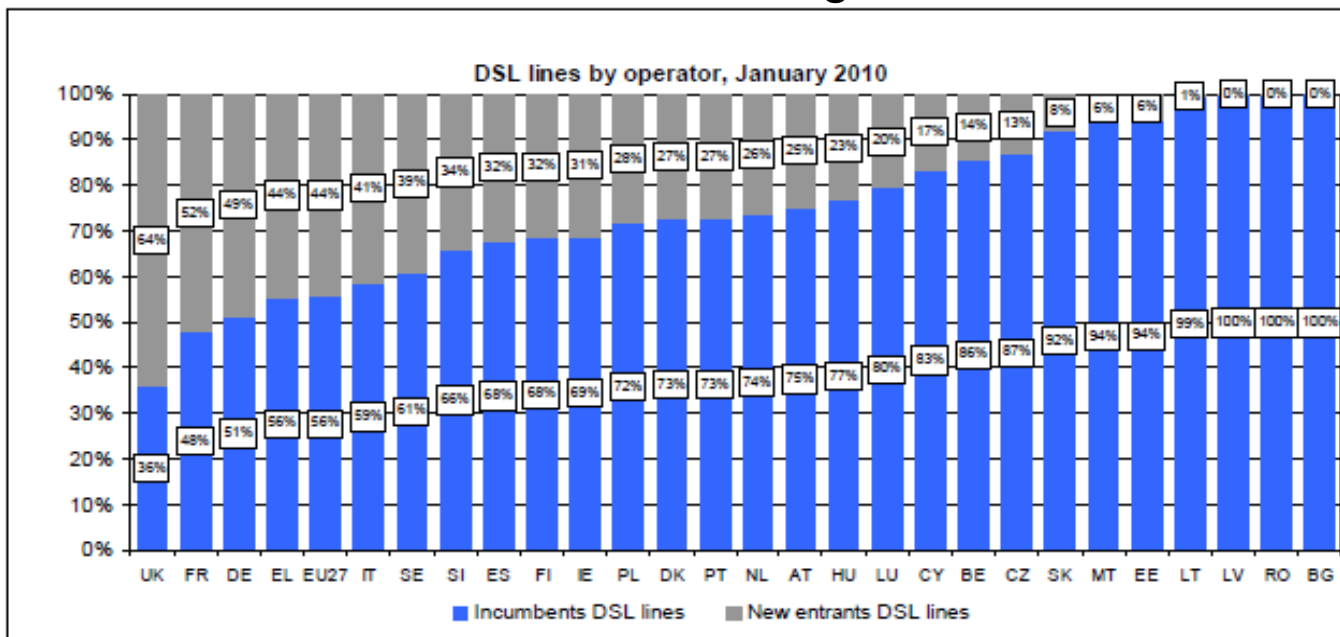
- Different market structure
- Different regulatory framework
- Different applicability of competition law

Market structure: US

- Most US homes could receive broadband from either a cable television provider or a telecommunications provider.
- Cable broadband reaches 38% of US households, ADSL and FTTH 30% (US FCC 2010 data).
- Competitive providers (using LLU, shared access, or bitstream) have largely disappeared in the US (less than 2% share), resulting in a market environment that is essentially *duopolistic*.
- Mobile broadband is widespread, but serves primarily as an *economic complement* to fixed, not as a *substitute*.

Market structure: EU

- The European competitive environment is richer overall.
- Cable coverage is uneven among Member States.
- LLU, shared access and bitstream are generally effective, but there are substantial differences among the Member States.



Source: 15th Implementation Report, 2010

- *Telecommunication services* are subject to numerous regulatory obligations; *information services* are subject to few explicit obligations.
- This distinction historically enabled the FCC to avoid regulating the Internet core.
- During the George W. Bush years, the FCC classified broadband access when bundled with Internet service to be an information service.
 - Weakened or lifted procompetitive remedies, thus reversing the growth of retail competition for DSL lines.
 - Lifted non-discrimination obligations.

- The FCC issued an Open Internet ruling in December, 2010.
 - **Rule 1: Transparency:** A provider of broadband Internet access service must publicly disclose accurate information regarding its network management practices, performance, and commercial terms sufficient for consumers to make informed choices ...
 - **Rule 2: No Blocking:** A provider of fixed broadband Internet access service shall not block lawful content, applications, services, or non-harmful devices, subject to reasonable network management.
 - **Rule 3: No Unreasonable Discrimination:** A provider of fixed broadband Internet access service shall not unreasonably discriminate in transmitting lawful network traffic over a consumer's broadband Internet access service.
- The ruling thus imposes fewer burdens on mobile networks.
- The ruling has been aggressively challenged in the courts and in the House of Representatives.

Regulation: EU

- Provisions for competitive access (LLU, shared access and bitstream) continue to be generally effective.
- Revisions to the regulatory framework were enacted in 2009, with transposition required in 2011.
 - The ability of end users to access content, applications or services of their choice is now an explicit goal of European policy.
 - Providers of electronic communication services must inform end users of their practices in regard to traffic management, and provide end users with the right to change providers without penalty if they are dissatisfied with a change in these practices.
 - Empowerment of NRAs to impose, if necessary, minimum QoS obligations on an SMP operator.

Competition law: EU and US

- In Europe, competition law is generally viewed as an *ex post* complement to *ex ante* regulation.
- In the US, by contrast, competition law is largely pre-empted by sector-specific regulation as a result of a number of court rulings (Trinko, Goldwasser).

Observations: Net Neutrality in the US and the EU

- Different markets, different regulatory arrangements.
- Network neutrality rules may be required in the US because a procompetitive regulatory policy was abandoned.
- Very little seems to be needed in Europe, because markets are sufficiently competitive to make harmful violations unlikely.
- „Trigger happy“ rules could do more harm than good.

Content versus network transmission: A debate



Through the Looking Glass, and what Scott found there

... with apologies to John Tenniel and Charles Lutwidge Dodgson (Lewis Carroll).

Content versus network transmission: A debate

- A.T. Kearney, “A Viable Future Model for the Internet” (2010):
 - “[T]hose who benefit from higher traffic volumes are those who generate traffic (typically content sites) and those who consume it (typically end users).
 - Those who have to build and operate the networks required to carry these traffic volumes earn almost no revenue from the former group and are often locked into flat rate price schemes with the latter group, continually decreasing because of retail competition.”

Content versus network transmission: A debate

- In a recent white paper, the European incumbents (ETNO) said:
 - “ETNO is not asking for increased regulatory intervention but aims to establish a reference for commercial negotiations.”
 - “ETNO [seeks a new model for interconnection that would] acknowledge the challenges of the new Internet economy and the principles that fair compensation is received for carried traffic and operators’ revenues should not be disconnected from the investment needs caused by rapid Internet traffic growth.”
 - “The goals for the new interconnection models are: enable incremental revenues by end-to-end QoS pricing and content value pricing; ...”

Network Operators and Content Providers: Who Should Bear the Cost?

- This debate has deep roots:

“The chief executive of AT&T, Edward Whitacre, told *Business Week* last year that his company (then called SBC Communications) wanted some way to charge major Internet concerns like Google and Vonage for the bandwidth they use. “What they would like to do is use my pipes free, but I ain't going to let them do that because we have spent this capital and we have to have a return on it,” he said.”

NY Times, March 8, 2006

Content versus network transmission: A debate

- How rapid is Internet traffic growth?
- How are costs growing over time?
- Do flat rate price schemes for those who consume content imply that prices necessarily decrease over time?



The perils of rapid growth in Internet traffic

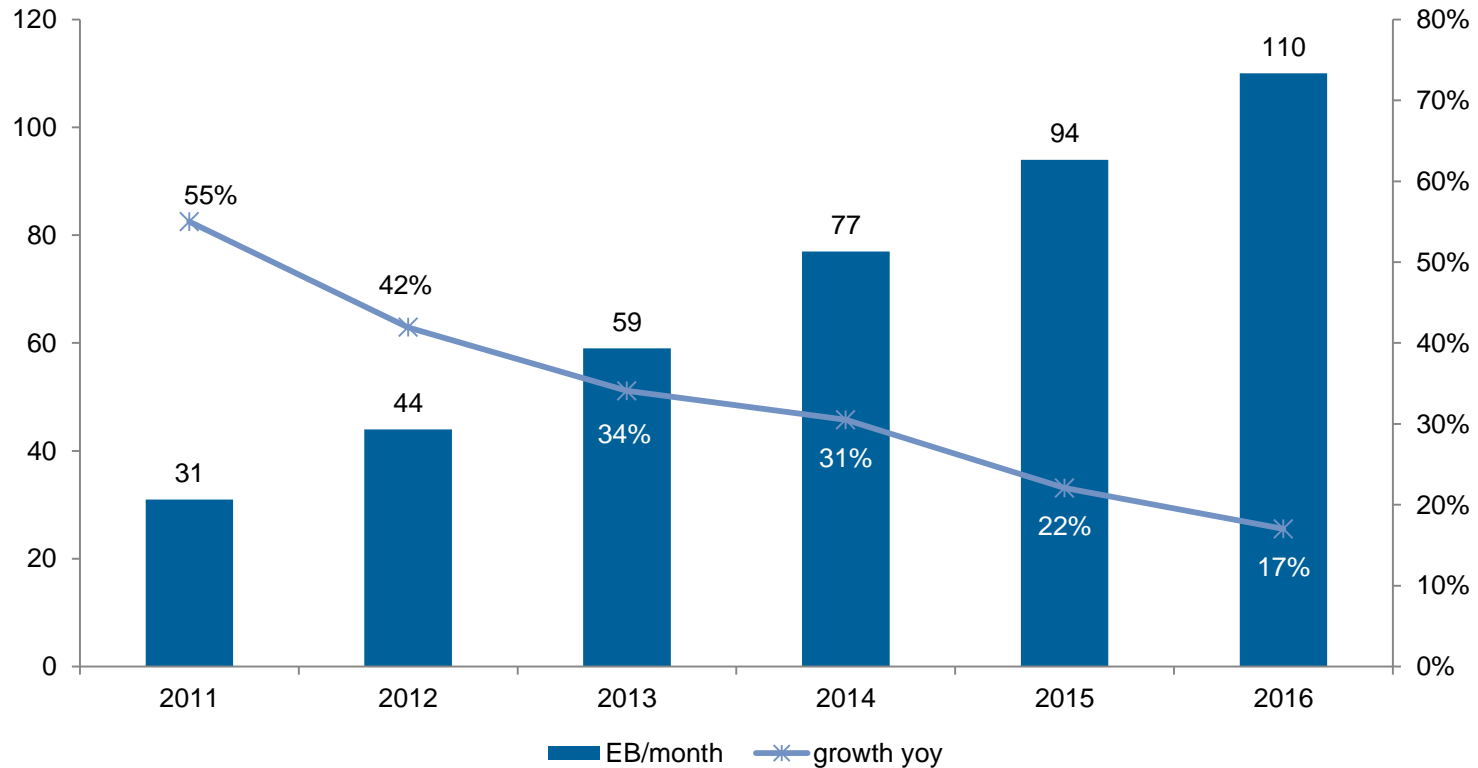
... with apologies to John Tenniel and Charles Lutwidge Dodgson.

Rapid growth in Internet traffic?

- Traffic volumes for Internet Protocol (IP) traffic are increasing, both for fixed and for mobile networks, but the percentage rate of increase is declining over time.
- The number of fixed broadband subscribers continues to increase, as does the number of mobile users who use data services.
- Traffic growth is largely a function of:
 - an increase in the number of subscribers, and
 - an increase in traffic per subscriber.
- Some costs are largely driven by the number of subscribers, and are largely independent of usage per subscriber.

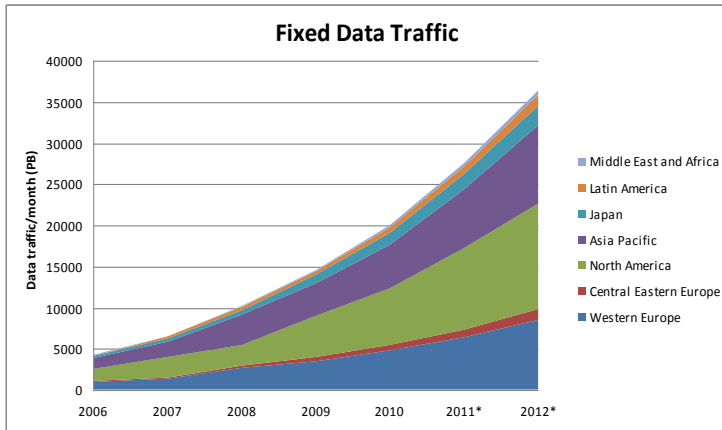
Rapid growth in Internet traffic?

- Traffic continues to grow, but the global rate of growth in percentage terms is declining over time.

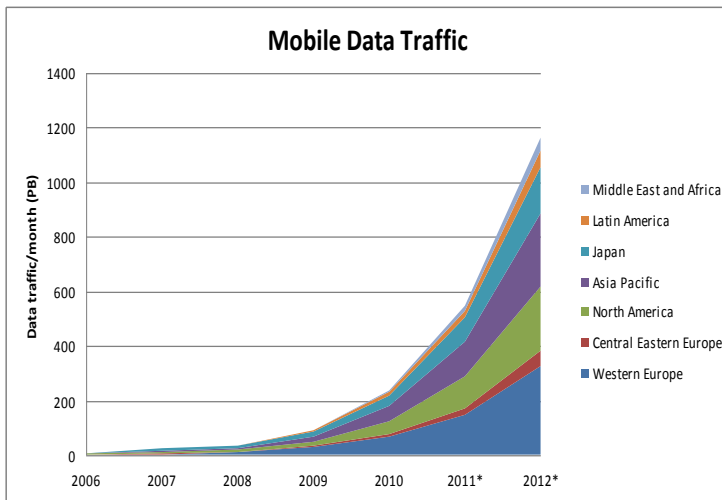


Source: Cisco (2012), WIK calculations.

Rapid growth in Internet traffic?



Traffic is indeed increasing in both the fixed and the mobile networks.



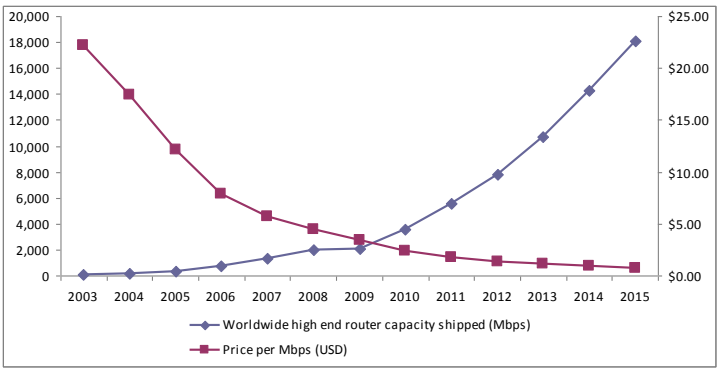
Source: Cisco (2011), WIK calculations.

Internet traffic growth drives increased investment needs

... with apologies to John
Tenniel
and Charles Lutwidge Dodgson.

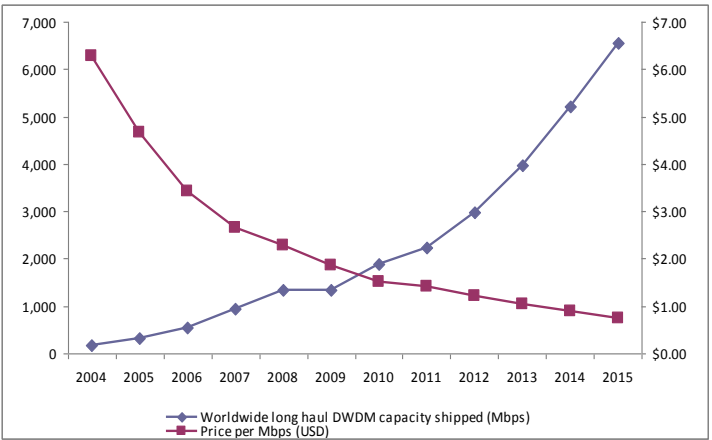


Rapid Internet traffic growth drives investment needs



Here we have the shipment quantities in Mbps and the price per Mbps (USD) for high end routers and for long haul DWDM optoelectronic equipment.

These are among the key cost drivers for Internet core and aggregation networks.



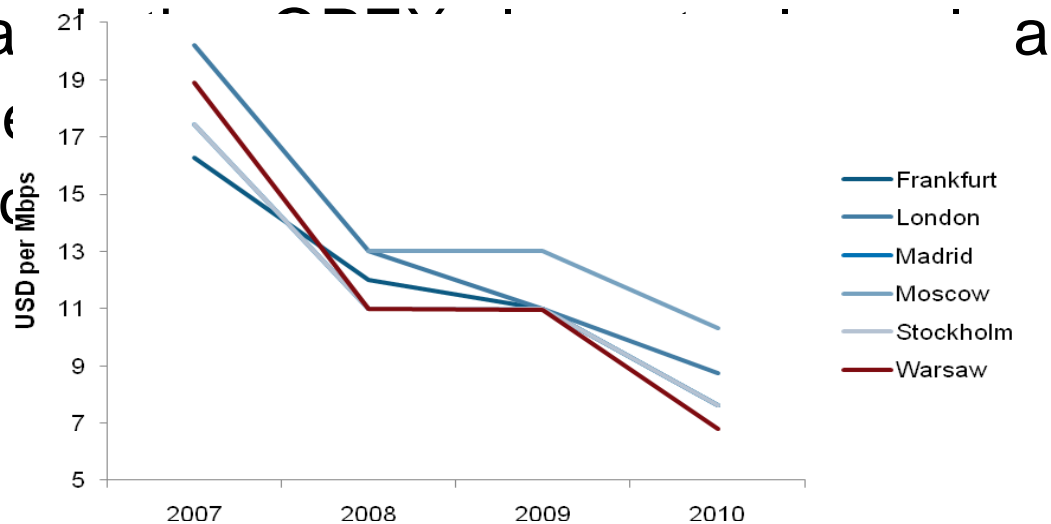
The growth in shipments generally tracks the Cisco projections.

The growth in *shipment volume* does not equate to a growth in *costs*, because the decline in unit costs is nearly in balance with it.

Source: Dell'Oro (2011), WIK calculations.

Rapid Internet traffic growth drives investment needs

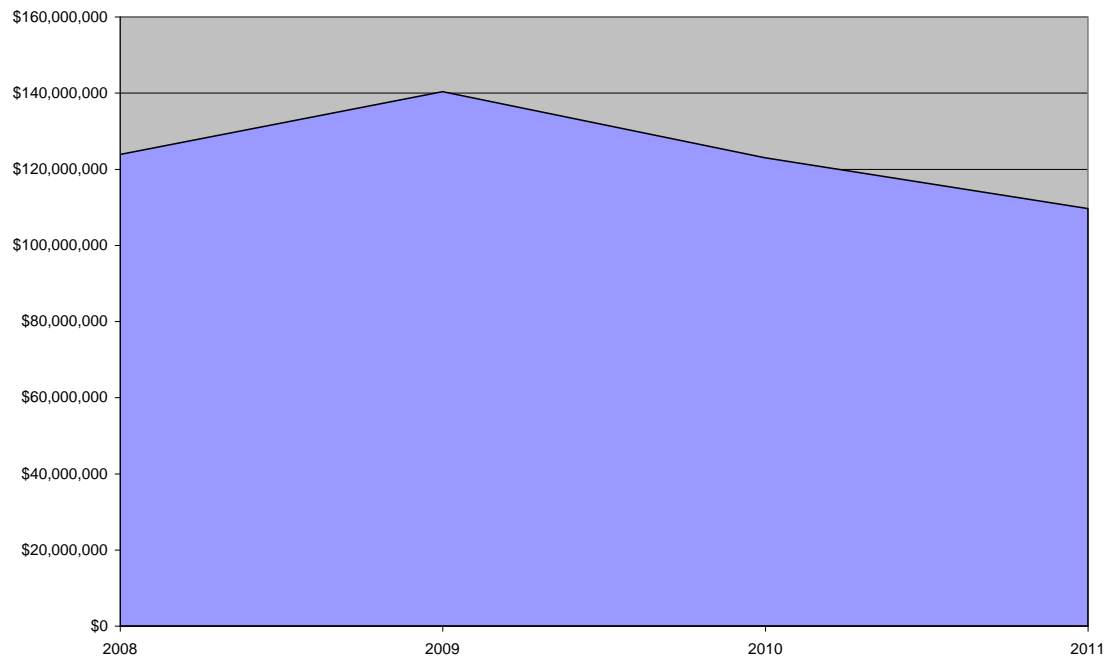
- Meanwhile, unit prices for global transit are declining rapidly.
- This decline reflects not only equipment costs but also circuits (over land and under water).
- Labour a small role number



Source: Telegeography (2011), WIK calculations.

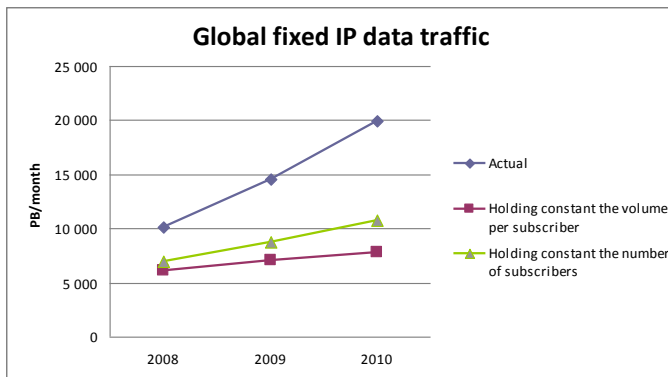
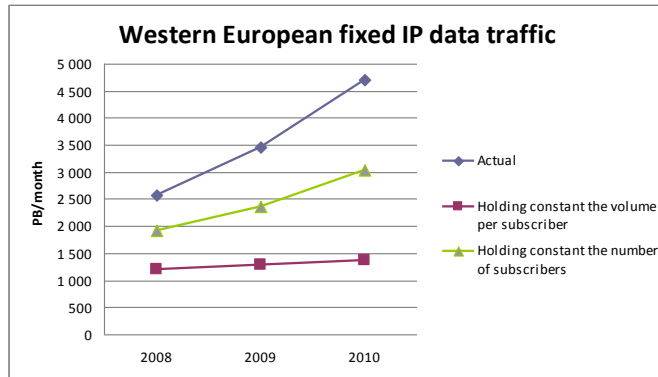
Rapid growth in investment needs driven by Internet traffic growth

- The monthly cost of carrying every bit of Western European Internet traffic, including growth, is flat or declining (and small in any case).
- Any self-supply is presumably cheaper than buying transit.



Source: Cisco (2011), Telegeography (2011), WIK calculations.

Rapid Internet traffic growth drives investment needs



Internet traffic growth is partly a response to increased use of Internet applications and content, and partly a result of increase in the number of subscribers.

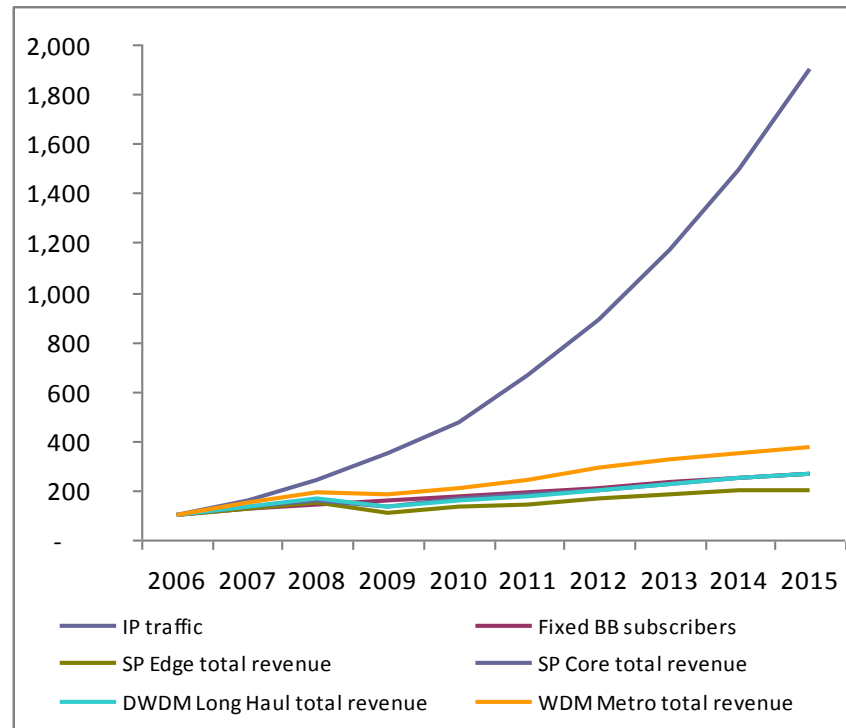
It is possible to distinguish between these two effects.

The effects on network operator profitability can be quite different. Increases in the number of subscribers equate to increased revenues.

Source: Cisco (2011), OECD (2011) WIK calculations.

Rapid Internet traffic growth drives investment needs

- This is consistent with the trend in underlying equipment costs, which tracks with subscribership and revenue, not with the volume of traffic.

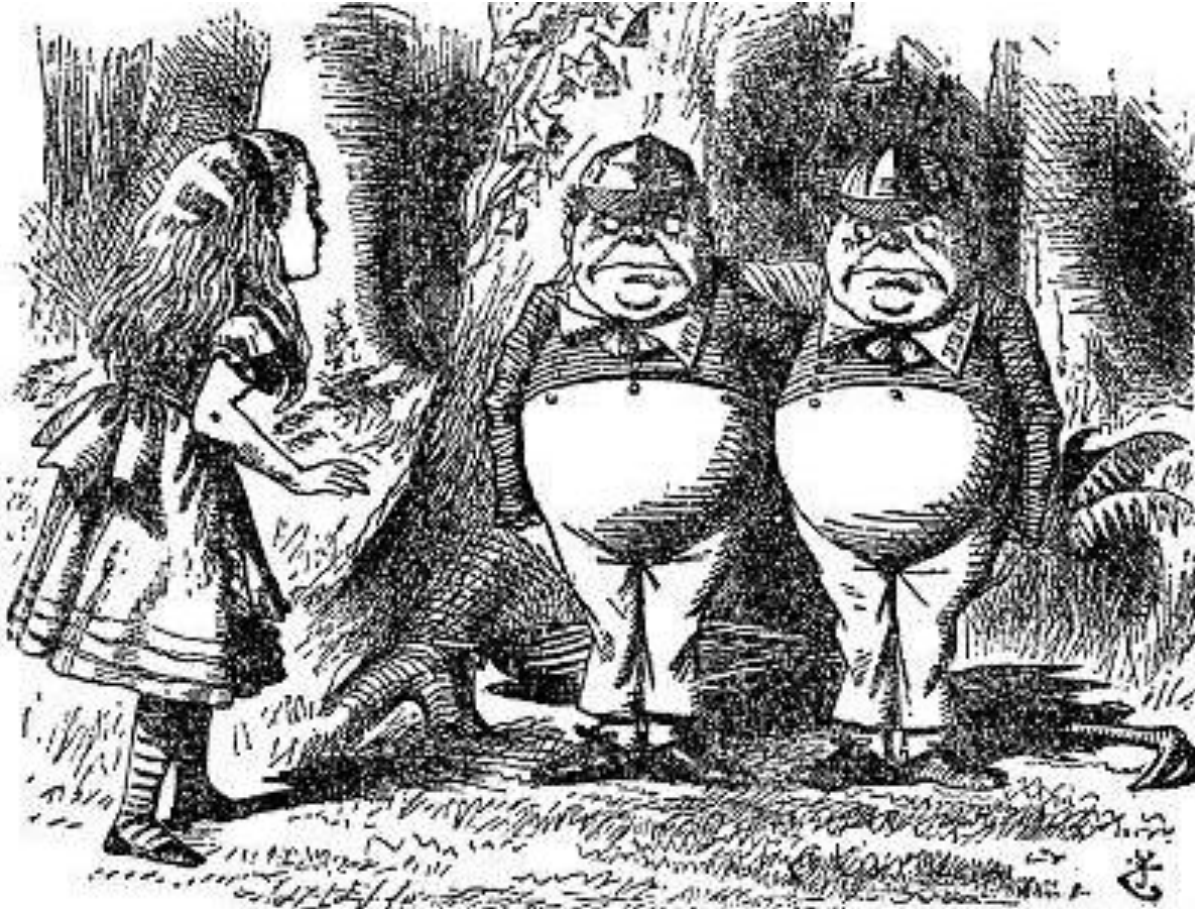


Source: Dell'Oro (2011), Cisco (2011), WIK calculations.

BEREC findings

- “[Content and application providers] make substantial payments for hosting and connectivity. Furthermore they pay for CDN services that bring content closer to the [consumer]. Therefore different from what is sometimes alleged by some telcos in the Net Neutrality debate there seems to be no free-riding problem. This holds even for those [content providers] that stick to their core activity which is the provision of content and/or applications without further vertically integrating along the value chain. Therefore BEREC conjectures that everything is covered and paid for in the Internet value chain (from content providers to the [users]).”

- BEREC consultation

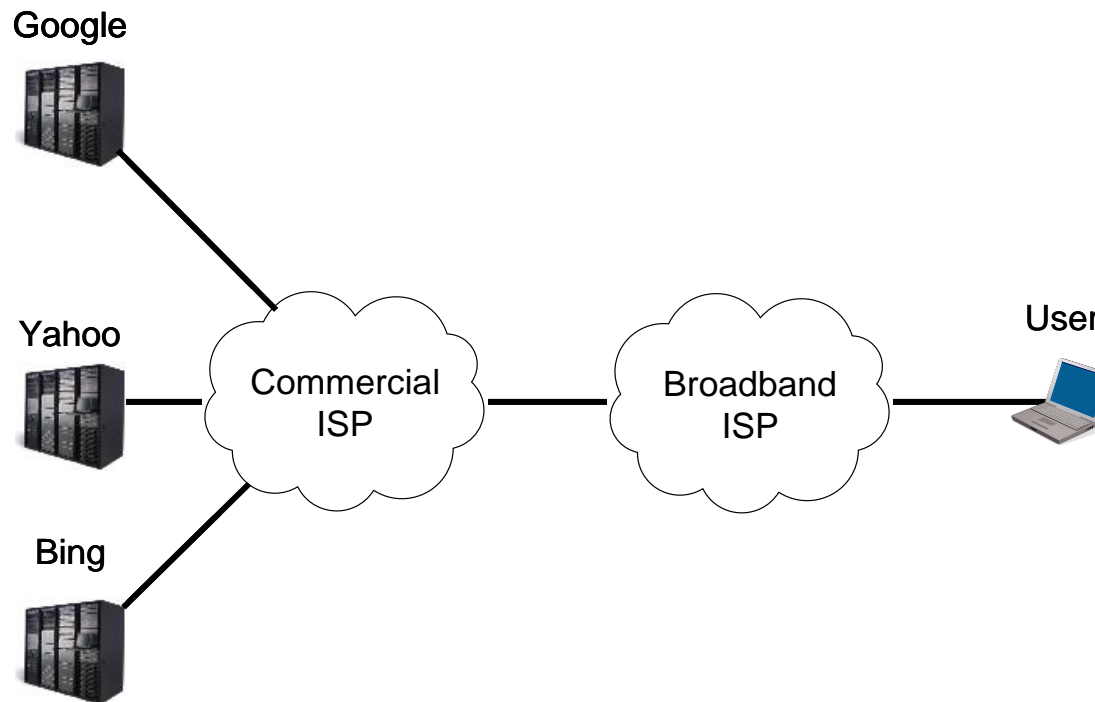


Two-sided markets

... with apologies to John Tenniel and Charles Lutwidge Dodgson.

Two-sided markets

- Recall that the objective in a two-sided market is to enhance societal welfare by maximizing participation and usage externalities.



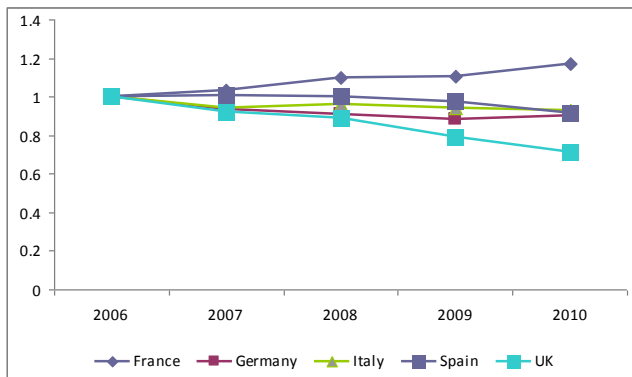
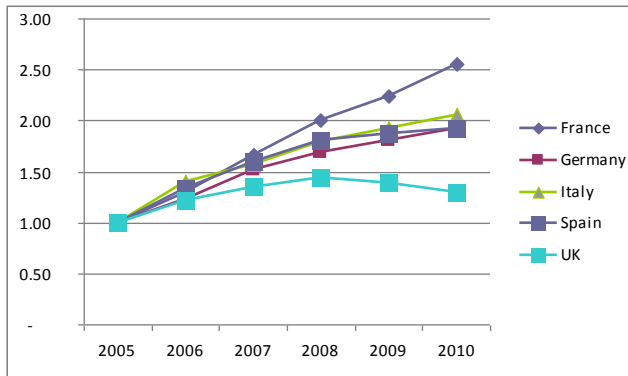
Two-sided markets: Price signals

- For many reasons, prices signals *should* be stronger to consumers than to content producers.
- Cost causation in the Internet is far more complex than in the traditional telephone network, but most decisions to consume content are made by the consumer.
- It is the consumer who decides to download a video – the content provider merely makes the content available.
- It is therefore entirely appropriate that the consumer be subject to stronger price signals than the content provider.

Two-sided markets: Price signals

- A large fraction of the cost of data transmission and receipt is associated with last mile access.
- The consumer decides how fast a broadband access is desired, and pays accordingly.
- The same is true of the content provider, but access benefits from huge economies of scale. Large content providers benefit from economies of scale in nearly all aspects of their operation.
- Once the access facility is in place, costs are largely independent of the volume of data. Truly usage-dependent costs typically represent a small fraction of the total costs of data transmission and receipt.
- For consumers to pay higher unit prices for data transmission than large content providers is not an anomaly. It is an expected and desirable result.

Two-sided markets and consumer WTP



Total fixed broadband subscriber revenue is increasing at a rate that reflects the growth in subscribership.

Fixed broadband subscriber revenue per subscriber (ARPU) is fairly steady.

The retail unit price is stable because underlying costs are stable.

Prices seem to move in both directions.

The end of unlimited mobile flat rates in the US demonstrates that prices can increase, even under a flat rate.

Source: IDATE data (2011), WIK calculations.

Two-sided markets and NGA

- Through the EU 2020 strategy, Europe seeks to achieve availability of 30 Mbps broadband to all Europeans by 2020, with half of all broadband consumers served at speeds of 100 Mbps or more.
- Consumers have only limited interest in NGA at present – incremental willingness to pay for ultra-fast broadband is only about € 5 per month, which is nowhere near enough to fund the initial investment needed in most parts of the national territory.
- Most estimates of the investment needed are in the neighbourhood of € 200 – 300 billion.

Network Type	Cost per home accessed [in €]					
	DE	FR	SE	PT	ES	IT
VDSL	457	n.v.	352	218	254	433
PON	2,039	1,580	1,238	1,411	1,771	1,110
P2P	2,111	2,025	1,333	1,548	1,882	1,160

Source: WIK (2008).

Two-sided markets and NGA

- Determining the right balance of payments in a two-sided market is complex. Many factors would need to be considered, including externalities and demand elasticities.
- The focus for NGA deployment and adoption to date has been on the *supply* side, not on the *demand* side.
- If one were going to take a two-sided market approach to NGA deployment, the optimal flow of payments could just as well be from network operators to content providers (i.e. in the opposite direction from that which has been suggested).
- If consumers are not convinced that ultra-fast connectivity is worth what it would cost, there is a clear need for more high value high bandwidth content.
- Without rigorous analysis, the answer is indeterminate.

Observations: Network Operators and Content Providers

- Total fixed network costs are increasing, as usual.
- The rate of total fixed network costs appears to be in balance with the rate of increase in subscribers, for at least some network operators. This means that the cost per user is relatively stable.
- Total retail revenue for fixed broadband has increased in proportion to the number of subscribers. Traffic growth driven by an increase in the number of subscribers should therefore raise no concerns.
- If costs had increased markedly, prices would have increased. Retail prices tend to move up or down in response to underlying costs, in this market as in most healthy competitive markets.

Observations: Network Operators and Content Providers

- The growth in mobile Internet traffic is quite stunning.
- Costs will likely increase near term (e.g. fibre to cell sites).
- All indications are that mobile network operators can and will find ways to adjust their retail prices to keep them in balance with their costs, provided that regulators do not prevent them from doing so.
- Price increases for mobile data should not necessarily be viewed as indicating market failure.

Observations: Network Operators and Content Providers

- “Generally, QoS classes at the network layer can be welfare enhancing as long as [consumers] can make an informed decision. This requires transparency.”
 - BEREC consultation
- In both the fixed and mobile networks, it is important that regulation not interfere with non-discriminatory QoS-aware commercial arrangements among willing parties.

- In conventional networks:
 - TRs would be very high absent regulation.
 - Regulated TRs should approximate long run incremental costs.
 - Those costs are declining rapidly due to IP-based technology.
- In Internet networks:
 - Transit prices are generally set by the market.
 - Peering charges, if any, are set by commercial negotiations.
 - Probably the best way to do it (but not everyone agrees).
- As traditional voice moves to IP:
 - IP-based voice termination cannot have a substantially different cost from that of traditional voice.
 - Basing the TR on the IP-based cost seems to make sense.
 - Charges for PSTN-to-IP conversion could make sense.

References

- Rochet, Jean-Charles/ Tirole, Jean (2004): Two Sided Markets : An Overview, March 2004, available at: http://faculty.haas.berkeley.edu/hermalin/rochet_tirole.pdf
- Marcus and Monti, Network Operators and Content Providers: Who Bears the Cost?" (September 13, 2011). Available at: <http://ssrn.com/abstract=1926768>
- BEREC, An assessment of IP-interconnection in the context of Net Neutrality: Draft report for public consultation, BoR (12) 33, 29 May 2012.
- Marcus et al., "Network Neutrality", a study for the European Parliament, 2011.
- Marcus et al., Interconnection in Next Generation Networks (NGNs), study for OSIPTEL (Peru), 2009.
- Marcus, Dieter Elixmann, Kenneth R. Carter, and numerous senior experts, *The Future of IP Interconnection: Technical, Economic, and Public Policy Aspects*, a study prepared for the European Commission, available at: http://ec.europa.eu/information_society/policy/ecomm/doc/library/ext_studies/future_ip_intercon/ip_intercon_study_final.pdf.
- Marcus, "IP-based NGNs and Interconnection: The Debate in Europe", *Communications & Strategies*, November 2008, http://www.idate.org/fic/revue_telech/831/CS72_MARCUS.pdf.
- Marcus, Dieter Elixmann, "Regulatory Approaches to Next Generation Networks (NGNs): An International Comparison", *Communications and Strategies* number 69, first quarter 2008. Preliminary versions are available on the web sites of TPRC 2007, and of the Center for Communication Law & Policy (CCLP) at USC, at http://cclp.usc.edu/assets/docs/RegulatoryApproachestoNGNs_AnInternationalComparison.pdf and <http://web.si.umich.edu/tprc/papers/2007/800/Regulatory%20Approaches%20to%20NGNs%20v1.2%20FINAL.pdf>.



See www.scottmarcus.eu, click on "publications".



wik-Consult GmbH
Postfach 2000
53588 Bad Honnef
Tel 02224-9225-0
Fax 02224-9225-68
eMail info@wik-consult.com
www.wik-consult.com