QoS and Interconnection in an NGN Environment

J. Scott Marcus, Senior Consultant

ITU Global Seminar on Quality of Service and Consumer Protection

Geneva - August 30, 2006

wik CONSULT

QoS and Interconnection in an NGN Environment

- Introduction
- Why differentiated QoS?
  - Application requirements
  - Technical considerations
- IP-based interconnection
- Challenges to inter-provider QoS deployment
  - Network externalities, transaction costs
  - Billing and accounting
- NGN interconnection, market power, and network neutrality
- Concluding remarks

wik CONSULT

Introduction

- IP-based NGNs represent the “marriage” of the Public Switched Telephone Network (PSTN) with the world of the Internet
- Very different interconnection arrangements prevail in these two worlds.
  - Different technology.
  - Different regulatory history.
  - Different industry structure.
- What should happen “when worlds collide?”

Introduction: Why do we regulate?

- Three potential reasons for *ex ante* regulation in general:
  - Market failures: Market power
  - Market failures: Desirable capabilities that would not deploy without help (some of which constitute “public goods”)
  - Manage limited resources (spectrum, numbers)
- At least the first two are relevant to NGN interconnection.

QoS – Application requirements

- The Internet was designed to operate correctly over any Data Link Layer technology, or combination of technologies; however, performance will vary based on the underlying technologies.
  - Bandwidth
  - Delay / Jitter
  - Packet loss
- This approach works wonderfully for delay-insensitive applications such as e-mail.
- It works well enough for interactive applications such as web browsing.
- It can pose challenges for real time applications, notably for bidirectional voice.

QoS – Application requirements

- For real time services such as bidirectional voice telephony traffic, it is important to avoid high mean delay and variability of delay.
  - Delay in excess of about 150 milliseconds causes “collisions”.
  - Buffering can address variability as long as the mean and variance are not too great.
  - The buffer then represents a fixed increment to the propagation delay.
- For streamed audio and video, delay is not necessarily a problem.
  - Initial delay of a few seconds is often sufficient to enable jitter buffers to deal with any subsequent variability.
  - May be perfectly acceptable for non-linear content.
  - Arguably less acceptable for “channel surfing” of linear content.

QoS – technical considerations

- A series of initiatives to implement inter-provider QoS on an inter-provider basis have generated scant results.
  - Early Nineties – RSVP
  - Late Nineties - DiffServ
  - Backbone peering attempts – especially circa 2000 – 2001
- Differentiated QoS *within a single network* is straightforward, and is widely implemented.

QoS – technical considerations

- Early Nineties – Integrated Services Architecture (RSVP)
  - Comprehensive system of prioritized delivery.
  - Required significant “soft state” in the routers.
- Perceived as too difficult to deploy.
- Not entirely true – BBN (my former employer) had working RSVP-capable production networks from the mid-Nineties.
  - Technical success.
  - Market failure.
  - No customer willingness to pay a significant premium for on-net differentiated service.

QoS – technical considerations

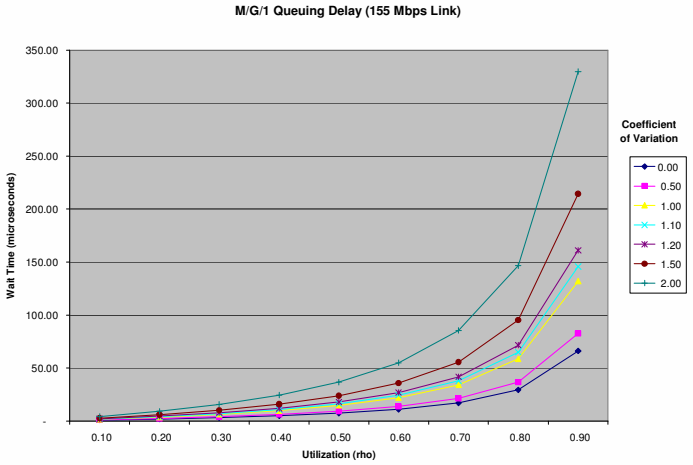
- Traffic is classified on entry to a network
  - Metered
  - Marked
  - Policed
  - Shaped
- Services implemented based on defined Per-Hop Behaviors (PHBs)
  - Queue processing (prioritization)
  - Queue management (drops)

QoS – technical considerations

- 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 company, BBN, had working production RSVP-compliant networks in 1995!
- QoS is widely employed *within* individual networks.
- In terms of the basic economics, QoS is not fundamentally hard.
- Nonetheless, there is no significant cross-provider roll-out to date.

➢ ***WHY NOT?***

QoS – technical considerations



M/G/1 queueing analysis of link performance

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

QoS – technical considerations

- For real time services such as voice telephony traffic, it is important that mean delay and variability of delay be held to values of not more than about 150 milliseconds in order to avoid “collisions”.
- For circuit speeds of T-3 (45 Mbps) and up, queuing delays in a properly designed network will generally be less than 1 millisecond per hop under normal operating conditions.
- Propagation delay (speed of light) will thus tend to dominate any variable queuing delays under normal operating conditions.

QoS – technical considerations

- IMPLICATION: Most of the time, and under normal conditions, variable delay in the core of the network(s) is unlikely to be perceptible to the VoIP user.
- 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 broadband Internet services)
  - When one or more circuits are saturated
  - When one or more components have failed
  - When a *force majeure* incident has occurred

QoS and Interconnection in an NGN Environment

- Introduction
- Why differentiated QoS?
  - Application requirements
  - Technical considerations
- IP-based interconnection
- Challenges to inter-provider QoS deployment
  - Network externalities, transaction costs
  - Billing and accounting
- NGN interconnection, market power, and network neutrality
- Concluding remarks

NGN interconnection – economic models

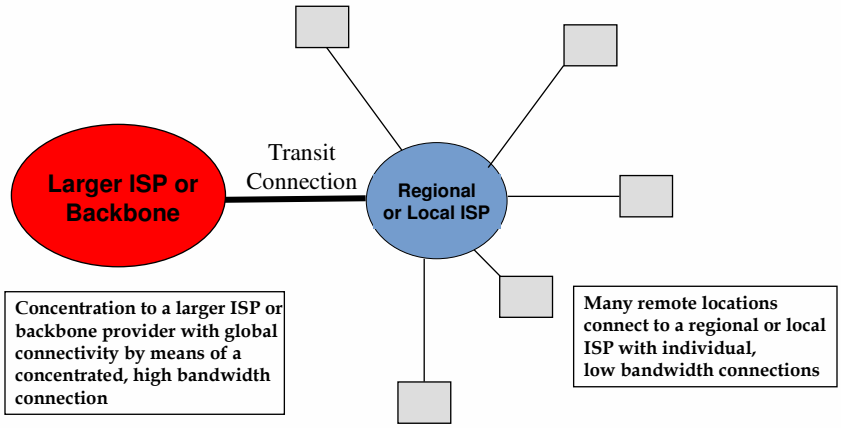
- An extensive economics literature exists about interconnection in the traditional PSTN world.
- An emerging literature deals with interconnection in the world of the Internet.
- We are in the early stages of understanding the relationships between the two.

NGN interconnection – peering and transit

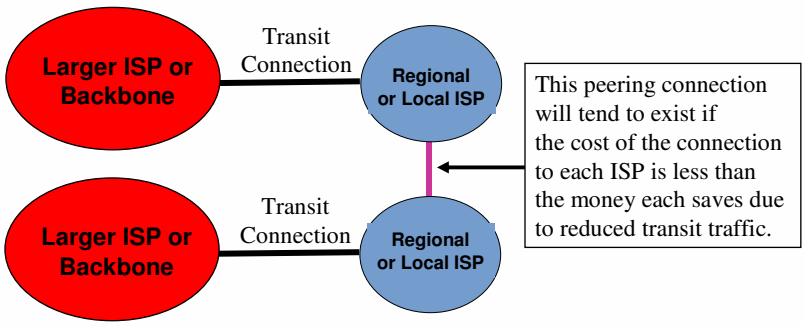
- *"Peering* is an agreement between ISPs to carry traffic for each other and for their respective customers. Peering does not include the obligation to carry traffic to third parties. Peering is usually a bilateral business and technical arrangement, where two providers agree to accept traffic from 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."

- NRIC V (advisory council to FCC)

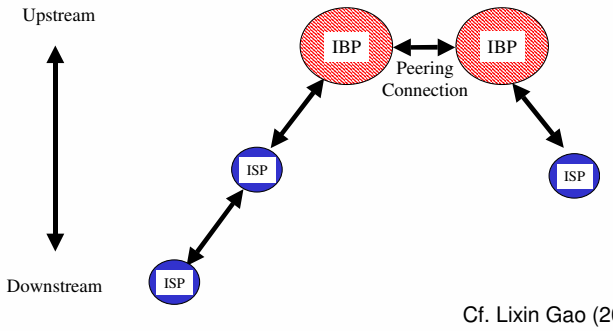
NGN interconnection – peering and transit



NGN interconnection – peering and transit



NGN interconnection – peering and transit

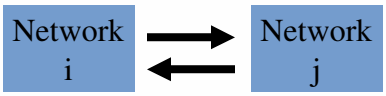


Cf. Lixin Gao (2000)

- In general, money flows upstream, while obligations flow downstream.
- Transit agreements are vastly simpler than peering agreements.
- In general, peering is a bilateral technical and commercial arrangement.

NGN interconnection – economic models

- Define:
  - c_o as cost of origination
  - c_t as cost of termination
  - a as an access charge levied on the sender
- Due to shortest exit, $c_t > c_o$
- Then
  - cost for the originating network is $c_o + a$
  - cost for the terminating network is $c_t - a$



The model extends in a straightforward way to accommodate multiple levels of quality of service (QoS).

Source: Laffont, Marcus, Rey and Tirole, "Internet Interconnection and the Off-Net-Cost Pricing Principle"

NGN interconnection – economic models

"A key difference with this telecommunications literature is that in the latter there is a missing price: receivers do not pay for receiving calls; … The missing price has … important implications:

… The operators' optimal usage price reflects their *perceived* marginal cost. But when operators do not charge their customers … for the traffic they receive, operator *i* 's perceived marginal cost of outgoing … traffic is … the unit cost of traffic is the on-net cost *c*, augmented by the expected off-net "markup". …

Comparing the two perceived marginal costs of outgoing traffic with and without receiver charge, for given access charge and market shares, the price for sending traffic is higher (lower) than in the presence of reception charges if and only if there is a termination discount (markup).

… In sum, the missing payment affects the backbones' perceived costs, and it reallocates costs between origination and reception."

Source: Laffont, Marcus, Rey and Tirole, "Internet Interconnection and the Off-Net-Cost Pricing Principle"

Challenges to inter-provider QoS deployment

- Some capabilities are worth vastly more, as more consumers adopt them. Nothing succeeds like success. This property is known as a *network effect* or a *network externality*.
- The societally optimal value of adoption of such services is not necessarily where the market would settle without "help".
- Different services have gotten past this *initial adoption hump* in different ways:
  - telephone - universal service
  - VCRs - widespread deployment for time shifting antedated the emergence of a rental industry.
  - CD players - vertical integration with recording studios
  - black and white television - industry / government standards

Cf. Rohlffs, *Bandwagon Effects in High-Technology Industries*, 2001.

Challenges to inter-provider QoS deployment

- Technical challenges, or economic challenges?
- Revenues
  - Limited customer willingness to pay a substantial premium.
  - Limited benefits until widely deployed (network effects).
- Costs
  - Agreements needed with many peering partners.
  - Economic *transaction costs* to negotiate each agreement.
  - Measurement, management and dispute resolution challenges.
- The business case is difficult to “prove in”.
- Implies difficulties in getting past the initial adoption hump.

QoS – billing and accounting challenges

- In an NGN world, the network service provider (the ISP) will not necessarily be the application service provider. A VoIP service or an IPTV provider will not necessarily be a network provider.
- The network provider will have only limited visibility into third party applications running over its network (and the user could further reduce visibility by encrypting the data).
- The unaffiliated application provider may have extensive visibility into the application that it provides, but only limited visibility into the use of network resources.
- Usage-based billing will be possible only to the extent that the usage can be rigorously and unambiguously measured.

QoS – billing and accounting challenges

- How will providers and customers ensure that service commitments are met? Whose statistics will govern?
- Competitive providers are reluctant to share statistics about their respective networks with one another, and peering agreements typically restrict the ability of the providers to disclose information about one another's networks to third parties. Can sufficient information be disclosed to customers?
- How will responsibility be allocated if a customer's traffic fails to achieve its committed service level specification? Traffic data can legitimately be interpreted in more than one way. Will it be possible to administer payments and penalties rigorously and fairly?
- How can providers prevent fraud? How can they distinguish between fraud and legitimate use?

QoS and Interconnection in an NGN Environment

- Introduction
- Why differentiated QoS?
  - Application requirements
  - Technical considerations
- IP-based interconnection
- Challenges to inter-provider QoS deployment
  - Network externalities, transaction costs
  - Billing and accounting
- NGN interconnection, market power, and network neutrality
- Concluding remarks

Market power and interconnection

- It is regulatory best practice to intervene in advance (*ex ante*) in order to enable competitive entry where an operator has Significant Market Power (SMP).
- Restrictions on interconnection have historically been one of the key means by which incumbents have sought to impede competitive entry.
- The migration to NGN will not necessarily eliminate SMP. Notably, market power associated with last mile bottlenecks will continue to be a significant regulatory concern for the foreseeable future.
- A new market power challenge has appeared, primarily in the U.S.: the *network neutrality* issue.

Market power and interconnection

- *Network neutrality* means different things to different people:
  - The possibility that an integrated ISP might offer better performance to some Internet sites than to others;
  - The possibility that an integrated ISP might assess a surcharge where a customer wants better-than-standard performance to certain Internet sites;
  - The fear that the integrated ISP might permit access only to affiliated sites, and block access to unaffiliated sites;
  - The fear that the integrated ISP might assess surcharges for the use of certain applications, or of certain devices;
  - The fear that the integrated ISP might disallow outright the use of certain applications, or of certain devices, especially where those applications or devices compete with services that the integrated ISP offers and for which it charges; and
  - The fear that the integrated ISP might erect “tollgates” in order to collect unwarranted charges from unaffiliated content providers who need to reach the integrated ISP’s customers.

Market power and interconnection

"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

Market power and interconnection

- Many of the concerns that have been raised in regard to network neutrality relate to behaviors that, in the absence of market power, would tend to *enhance* consumer welfare.
  - Some would appear to represent legitimate price discrimination.
  - Others enforce the economic property of *excludability* (the ability to prevent someone from using a service that he did not pay for) in support of price discrimination.
- The form of market power that could potentially be exploited in anticompetitive ways in connection with network neutrality relates to network externalities (where the value of a service depends on the number of users of the service). (Cf. Katz and Shapiro (1985)).

Market power and interconnection

- The degree to which this issue has heated up recently in the U.S. probably reflects several underlying root causes:
  - Increasing concentration of the retail broadband market, coupled with the continued decline of competitive market share in the wholesale DSL segment (currently less than 4%).
  - Effective elimination (Fall 2005) of obligations to treat content in a nondiscriminatory manner.
- The network neutrality debate is unlikely to manifest itself in the same way in Europe, where about 50% of all DSL is provided by new entrants.

Market power and interconnection

- Trying to address these network neutrality challenges through regulation *ex ante* (in advance) is likely to prove extremely difficult.
  - Difficult to distinguish between welfare-enhancing practices versus harmful anticompetitive practices.
  - Difficult to anticipate the tactics of market players.
- A first line of defense for regulatory authorities should instead be to maintain the competitiveness of the underlying markets, especially as regards broadband Internet access and as regards high capacity Internet transit. Service-based competition (rather than facilities-based competition) would be sufficient for this purpose.
- In countries where competition law provides an *ex post* (after the fact) complement to regulation, it might be more appropriate to deal with occasional or sporadic problems related to network neutrality through the exercise of competition law.

Concluding remarks

- Differentiated QoS is important for certain applications, especially bidirectional real time voice and video.
- Consumer willingness to pay a premium for QoS is likely to be less than many operators assume, because best-efforts services will work well enough in most environments most of the time (unless intentionally degraded).
- IP-based NGNs are likely to support QoS internally, but QoS *between* providers is likely to face ongoing challenges due to a mix of factors related to network externalities and transaction costs.
- Government sometimes plays a role in getting new services past the initial adoption hump. In this case, it is unclear what government initiatives, if any, might foster deployment and adoption. Beyond that, it is not clear that intervention is warranted.

Concluding remarks

- The migration to IP-based NGNs risks will not in and of itself eliminate all forms of market power. Moreover, the migration may introduce new forms of market power.
- In a competitive market, QoS and second order price discrimination will tend to be welfare enhancing.
- The network neutrality debate in the US reflects the hazards of rapid withdrawal of regulation in a market where concentration is increasing over time.
- Policymakers should continue to focus significant attention on ensuring competitive markets, especially for key inputs such as (1) residential broadband and (2) leased lines.

Papers

"Interconnection in an NGN Environment", a background paper commissioned for the ITU New Initiatives Programme workshop on "What rules for IP-enabled Next Generation Networks?" held on 23-24 March 2006 at ITU Headquarters, Geneva. Available at: http://www.itu.int/osg/spu/ngn/documents/Papers/Marcus-060323-Fin-v2.1.pdf. Also available as WIK Discussion Paper 274 (see http://www.wik.org/content_e/diskus/274.htm).

"WIK Workshop on 'Bill and Keep' Interconnection Arrangements", a summary of WIK's ground-breaking April 2006 workshop, in WIK's newsletter number 63, April 2006, pages 11-15, available *in English* at: http://www.wik.org/content/newsletter/nr63.pdf. The presentations from the workshop are available at: http://www.wik.org/content/bill_keep/konf_billkeep_main.htm.

"Framework for Interconnection of IP-Based Networks -- Accounting Systems and Interconnection Regimes in the USA and the UK", a background paper prepared for the German Federal Network Agency's study group on a Framework for Interconnection of IP-Based Networks, 27 March 2006, available at: http://www.bundesnetzagentur.de/media/archive/6201.pdf.

With Justus Haucap, "Why Regulate? Lessons from New Zealand", *IEEE Communications Magazine*, November 2005, available at: http://www.comsoc.org/ci1/Public/2005/nov/ (click on "Regulatory and Policy").

"The Challenge of Telephone Call Termination Fees", *Enterprise Europe*, January 2005. Available at: http://www.european-enterprise.org/public/docs/EEJ.pdf.

"Call Termination Fees: The U.S. 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.

"Evolving Core Capabilities of the Internet", *Journal on Telecommunications and High Technology Law*, 2004, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=921903.

With Jean-Jacques Laffont (deceased), 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, available at http://www.rje.org/abstracts/abstracts/2003/rje.sum03.Laffont.pdf. An earlier version of the paper appeared as "Internet Peering", *American Economics Review*, Volume 91, Number 2, May 2001.

wik CONSULT

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